November 2012

Crowdsourced Geospatial Data

A report on the emerging phenomena of crowdsourced and user-generated geospatial data

Matthew T. Rice, Fabiana I. Paez, Aaron P. Mulhollen, Brandon M. Shore

Department of Geography and Geoinformation Science
George Mason University
4400 University Drive
Fairfax, VA 22030-4444

Douglas R. Caldwell

Topographic Engineering Center
U.S. Army Engineer Research and Development Center
7701 Telegraph Rd.
Alexandria, VA 22315-3864

Annual Report, BAA #AA10-4733, Contract #W9132V-11-P-0011

Prepared for

U.S. Army Topographic Engineering Center
U.S. Army Engineer Research & Development Center
U.S. Army Corps of Engineers

Under Work Unit 633734T0800

Monitored by

U.S. Army Topographic Engineering Center
7701 Telegraph Road, Alexandria, VA 22315-3864

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
29-11-2012

2. REPORT TYPE
Annual

3. DATES COVERED *(From - To)*
Jun-2011 - Aug-2012

4. TITLE AND SUBTITLE
Crowdsourced Geospatial Data
A report on the emerging phenomena of crowdsourced and user-generated geospatial data

5a. CONTRACT NUMBER
W9132V-11-P-0011

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Rice, Matthew T.
Caldwell, Douglas R.
Paez, Fabiana I.
Mulhollen, Aaron P.
Shore, Brandon M.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
633734T0800

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
George Mason University
4400 University Drive
Fairfax, VA 22030-4444

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army Engineer Research and Development Center (ERDC)
Topographic Engineering Center (TEC)
7701 Telegraph Road
Alexandria, VA 22135-3864

10. SPONSOR/MONITOR'S ACRONYM(S)
ERDC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Crowdsourced geospatial data (CGD) is an important emerging trend that will influence future methods for geospatial data production and use. Related to broader developments in user-generated content, CGD involves the participation of end-users, many of who are untrained in the geospatial sciences but have a high degree of interest in geospatial technology. Working collectively, these end-users collect, edit, and produce datasets; create mapping applications, and develop tools for CGD. Crowdsourced geospatial data production is typically an open, lightly controlled process with few constraints, specifications, or quality assurance processes. This sharply contrasts with the less flexible and more controlled authoritative geospatial data production practices of national mapping agencies and businesses. Adoption of CGD and production methods has been a concern, especially to Government organizations, due quality concerns related to differences in production methods.
We review CGD projects addressing common geospatial data collection tasks and demonstrating varied approaches to quality control, including hybrid projects that mix crowdsourced geospatial data and tools with authoritative data. The most common methods for quality assessment are summarized along with a comprehensive set of fitness-for-use considerations. Based on this information, lessons learned and future trends are summarized.

15. SUBJECT TERMS
crowdsourcing, geospatial data, user-generated content, quality assessment

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	U	147	19b. TELEPHONE NUMBER *(include area code)*
U	U	U			

Abstract:

Crowdsourced geospatial data (CGD) is an important emerging trend that will influence future methods for geospatial data production and use. Related to broader developments in user-generated content, CGD involves the participation of end-users, many of who are untrained in the geospatial sciences but have a high degree of interest in geospatial technology. Working collectively, these end-users collect, edit, and produce datasets; create mapping applications, and develop tools for CGD.

Crowdsourced geospatial data production is typically an open, lightly controlled process with few constraints, specifications, or quality assurance processes. This sharply contrasts with the less flexible and more controlled authoritative geospatial data production practices of national mapping agencies and businesses. Adoption of CGD and production methods has been a concern, especially to Government organizations, due quality concerns related to differences in production methods.

We review CGD projects addressing common geospatial data collection tasks and demonstrating varied approaches to quality control, including hybrid projects that mix crowdsourced geospatial data and tools with authoritative data. The most common methods for quality assessment are summarized along with a comprehensive set of fitness-for-use considerations. Based on this information, lessons learned and future trends are summarized.

Contents

Figures and Tables

Figures

Tables

Preface

This report is a deliverable product of Department of the Army Broad Agency Announcement (BAA) #AA10-4733, Contract #W9132V-11-P-0011, under Work Unit 633734T0800,“Data Level Enterprise Tools.” The study for this report was conducted jointly by ERDC technical lead Douglas R. Caldwell and George Mason University faculty and research staff, in support of the U.S. Army Engineering Research & Development Center. Report author Matthew T. Rice wishes to acknowledge the financial support of Dean Vikas Chandhoke, George Mason University College of Science during the preliminary phase of this research project, and Fabiana I. Paez, for her editorial support.

Unit Conversion Factors

Multiply	By	To Obtain
Feet	0.3048	meters
miles (nautical)	1,852	meters
miles (U.S. statute)	1,609.347	meters
miles per hour	0.44704	meters per second

1 Introduction to Crowdsourcing and Crowdsourced Geospatial Data

Background

In early December 2004, a group of 40 experts from academia, business, and government met in Santa Barbara, California to discuss strategic advancements in geographic information science. The focus of the meeting was on the emerging and changing information landscape associated with Web 2.0, social media, and distributed information sharing communities.

The final report from this meeting, summarizing the consensus research priorities of the expert group, suggested that a new approach to geographic information sharing was emerging, where distributed geographic data, services, and information would be shared over the computer networks.[1] The report suggests that tools would emerge to facilitate the sharing of distributed collections of data, information, and services. An important point of emphasis is the assertion that the most compelling application domain for this emerging trend would be in the area of natural disasters and emergency management. Any emerging trend in geospatial data management, data sharing, and data integration, would be effectively focused by best use scenarios in the domain of emergency services and disaster response, which would highlight the benefits of distributed information sharing networks and data integration using distributed collections of information.

In the years following this December 2004 meeting, a sequence of unfortunate natural disasters occurred which confirmed the predictions of the report authors, beginning with the devastating December 26th, 2004 Indian Ocean earthquake and tsunami that occurred just three weeks after the meeting. Hurricane Katrina (September 2005), the Wenchuan Earthquake (May 2009), the Santa Barbara wildfires (2007-2009), and the Haitian Earthquake (January 2010) focused international attention on the immediate need for maps and geospatial data of the impacted areas and the critical role of geographically-distributed information sharing communities in providing that information.

[1] Michael F. Goodchild et al., *Report of the NCGIA Specialist Meeting on Spatial Webs* (Santa Barbara, CA: NCGIA, April 2005).

[2] "OpenStreetMap Statistics," *OpenStreetMap*, n.d.,

An important related development during these natural disasters has been the emergence of a large body of end-users creating, contributing, editing, and displaying massive amounts of geospatial data outside the normal authoritative channels.

Project	Contributors	Contributions
OpenStreetMap[2]	Over 720,000	30,264,55,008 GPS Points 145351000 Ways
Old Weather[3]	Over 27,000	1,659,212 Weather Observations
Wikipedia	Over 17,000,000[4]	4,029,897 Content Pages[5]

Michael F. Goodchild has labeled this critical participation of the end-user community and the associated information sharing practices as *volunteered geographic information* (VGI).[6] Authors have also described this development as a form of crowdsourcing, therefore suggesting the descriptor *crowdsourced geospatial data* (CGD), a term that will be used in this report.[7] Zook et al.,[8] Goodchild et al.,[9] and Elwood et al.[10] provide detailed summaries and analysis of this emerging, geospatial crowdsourcing phenomenon, describing the trend as a "paradigmatic shift in how geographic information is created and shared."[11] This report presents the context for this paradigmatic shift, the relevant considerations and important

2 "OpenStreetMap Statistics," *OpenStreetMap*, n.d., http://www.openstreetmap.org/stats/data_stats.html

3 Phillip, "One Million, Six Hundred Thousand New Observations," Blog, *Old Weather Blog*, July 23, 2012, http://blog.oldweather.org/2012/07/23/one-million-six-hundred-thousand-new-observations/

4 "Wikipedia:Wikipedians," Encyclopedia, *Wikipedia, the Free Encyclopedia*, August 16, 2012, http://en.wikipedia.org/wiki/Wikipedia:Wikipedians#Number_of_editors

5 "Statistics," Encyclopedia, *Wikipedia, the Free Encyclopedia*, n.d., http://en.wikipedia.org/wiki/Special:Statistics

6 Michael F. Goodchild, "Citizens as Sensors: The World of Volunteered Geography," *GeoJournal* 69, no. 4 (2007): 211–221.

7 A. M Ruitton-Allinieu, "Crowdsourcing of Geoinformation: Data Quality and Possible Applications" (2011); Michael F. Goodchild and J. Alan Glennon, "Crowdsourcing Geographic Information for Disaster Response: a Research Frontier," *International Journal of Digital Earth* 3, no. 3 (September 2010): 231–241; Matthew Zook et al., "Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake," *World Medical & Health Policy* 2, no. 2 (July 21, 2010): 6–32.

8 Zook et al., "Volunteered Geographic Information and Crowdsourcing Disaster Relief."

9 Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

10 Sarah Elwood, Michael F. Goodchild, and Daniel Z. Sui, "Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice," *Annals of the Association of American Geographers* 102, no. 3 (May 2012): 571–590.

11 Ibid., p. 571.

facets of the shift, and its significance within geographic information systems and geospatial technology.

As a starting point, this report identifies and defines relevant terminology and information on the genesis of the CGD, and presents the wider context for this emerging trend. Subsequent sections of this report discuss CGD in the context of geographic information systems and authoritative data production systems, sources and examples of CGD, data quality considerations for CGD, evaluating fitness-for-use of CGD, significant trends and lessons learned from CGD-related projects, and a summary of CGD.

Definition of Terms Associated with Crowdsourced Geospatial Data

The growth of the Internet over the last two decades has fundamentally changed the way geospatial information is produced, stored, disseminated, and used, with a change from centralized production and dissemination to a more complex arrangement of traditional authoritative sources and end-users. Elwood et al.[12] suggests that these changes are related to a larger movement of user-generated content (UGC), as seen in familiar projects such as Wikipedia, where content is contributed and edited by a community of end-users. Crowdsourcing, a term used to describe this process of collective authorship by a community of end-users, can take a variety of forms within the geospatial domain, reflecting the primary types of crowdsourcing suggested by Howe.[13] Some projects involve elements of crowd wisdom; others involve crowd creation, crowd voting, and crowd funding. A primary focus of this report is crowd creation, where geospatial data is produced and contributed by end-users and described as CGD.

CGD is derived from non-authoritative sources consisting primarily of end-users participating in social media and Web 2.0 activities. CGD can be primarily geospatial in nature, or could simply be an associated geospatial characteristic of non-geospatial information. CGD can be asserted by the end-users, or could be the product of active harvesting and synthesis.[14]

[12] Ibid.

[13] Jeff Howe, *Crowdsourcing : why the power of the crowd is driving the future of business* (New York: Crown Business, 2008).

[14] Anthony Stefanidis, Andrew Crooks, and Jacek Radzikowski, "Harvesting Ambient Geospatial Information from Social Media Feeds," *GeoJournal* (December 4, 2011), http://www.springerlink.com/index/10.1007/s10708-011-9438-2

As with any emerging trend, terminology and standard reference language take time to develop and gain acceptance. Currently, several related terms are used to describe movements, practices, and characteristics related to CGD: VGI, as identified above, and ambient geographic information (AGI). Other related terms of interest include citizen science, participatory sensing, and neogeography; mentioned here in order to clarify the material presented in this report.

The 2007 paper, "Citizen Sensors: The World of Volunteered Geography" by Michael Goodchild remains the most highly cited work in the geospatial crowdsourcing and geospatial web domain. In this work, VGI is described as a Web 2.0 based movement where end-users contribute geographic information to augment and replace existing sources of information such as printed maps, remotely-sensed images, and other web content.[15]

Goodchild cited the general decline in availability of printed maps, updates to digital map documents, and software applications such as Google Earth as motivations for end-users to contribute geospatial information. He also noted the significant amount of geospatial content in applications such as Wikimapia,[16] which contained more than 4.2 million entries at the time he wrote the paper. This large volume of end-user generated geospatial content equaled the size of the Alexandria Digital Library's gazetteer, which contains a comprehensive worldwide coverage of geographic names and feature types, compiled from US government sources.

An important characteristic of VGI is that end-users assert the information, which therefore lacks the authoritative stamp of approval, certification, or quality assessment typically done by a governmental mapping organization. This does not mean, however, that the information is inaccurate or unreliable. Many authors have noted that a primary benefit of VGI is that it is often contributed by end-users with significant local geographic expertise. These end users, while lacking the formal training, structure and authority of a governmental mapping organization, may be more familiar with local geographic conditions and characteristics. Additionally, end-users are able to contribute local geographic information more often and faster than any governmental mapping organization, which may have regular, periodic update cycles.

[15] Goodchild, "Citizens as Sensors: The World of Volunteered Geography."

[16] "Wikimapia - Let's Describe the Whole World!," n.d., http://wikimapia.org/

AGI is information harvested from sensors, observations, and social media feeds. AGI represents the geographic associations and footprints of social media, or rather, the "momentary social hotspots" in the human landscape.[17] Web 2.0 platforms such as Twitter, Facebook, Flickr, and YouTube have large volumes of information with geographic footprints or geographic associations that can be used in geospatial analysis and synthesis. The original end-users contributing to these platforms may not have intended the information to be geographic or to have a specific geographic or geospatial purpose. In that sense, AGI differs from VGI. Yet it too may be viewed as an evolution and extension of geospatial data availability as CGD.

Other general terms of interest in this report include UGC, which can be thought of as any data, information, or material contributed by end users rather than by a centralized authority. Primary examples of UGC are Wikipedia and Facebook, both of which have vast collections of information generated by the end-user community. As the broadest term, UGC includes CGD as a specific component.

Another term of interest is 'citizen science', which is UGC with a specific scientific emphasis, and often the result of public engagement with experts in the area of data collection. A primary example of citizen science is the Christmas Bird Count,[18] an effort organized by the National Audubon Society each December to conduct a comprehensive bird and wildlife census with the help of local volunteers.

The term 'participatory sensing' has been used to describe a citizen science-related activity that uses the power of mobile computing and sensing devices to gather information. End-users with mobile computing devices and sensors form interactive, participatory sensor networks to gather, analyze, and share information.

A final term of interest for this report is 'neogeography', described by Turner,[19] Goodchild,[20] Rana et al.,[21] and others as the blurring or mixing

---

[17] Stefanidis, Crooks, and Radzikowski, "Harvesting Ambient Geospatial Information from Social Media Feeds."

[18] "Christmas Bird Count," *National Audubon Society Birds*, n.d., http://birds.audubon.org/christmas-bird-count

[19] Andrew Turner, *Introduction to Neogeography* (O'Reilly Media, Inc., 2006).

[20] Michael F. Goodchild, "NeoGeography and the Nature of Geographic Expertise," *Journal of Location Based Services* 3, no. 2 (2009): 82–96.

of distinctions between authoritative geospatial data producers and communicators, and the end user or consumer of geospatial information. It is often conceptualized as the involvement and participation of untrained end users in the formerly restricted domains associated with authoritative data producers and communicators.

This report will build on these definitions and background information to contextualize CGD in geographic information systems (GIS), provide examples and sources of CGD, report on the data quality of CGD, demonstrate the fitness-for-use of CGD, and show significant trends and lessons learned from CGD projects. Several excellent sources of information on CGD and related topics are contained in an appendix, with a select number of resources identified for further study and consideration. A recent publication that merits close attention along with this report is the edited volume by Daniel Sui, Sarah Elwood, and Michael Goodchild titled "Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice" (2013),[22] which contains a number of excellent overview chapters on topics such as CGD services,[23] and future prospects of CGD.[24]

---

[21] S. Rana and T. Joliveau, "Neogeography Phenomena-Some Thoughts on It's Beginning, Future and Related Issues" (n.d.).

[22] Daniel Sui, Sarah Elwood, and Michael F. Goodchild, eds., *Crowdsourcing Geographic Knowledge Volunteered Geographic Information (VGI) in Theory and Practice.* (New York, NY: Springer, 2013).

[23] Jim Thatcher, "From Volunteered Geographic Information to Volunteered Geographic Services," in *Crowdsourcing Geographic Knowledge*, ed. Daniel Sui, Sarah Elwood, and Michael Goodchild (Dordrecht: Springer Netherlands, 2013), 161–173, http://www.springerlink.com/index/10.1007/978-94-007-4587-2_10

[24] Sarah Elwood, Michael F. Goodchild, and Daniel Sui, "Prospects for VGI Research and the Emerging Fourth Paradigm," in *Crowdsourcing Geographic Knowledge*, ed. Daniel Sui, Sarah Elwood, and Michael Goodchild (Dordrecht: Springer Netherlands, 2013), 361–375, http://www.springerlink.com/index/10.1007/978-94-007-4587-2_20

2 Crowdsourced Geospatial Data Production versus Traditional Geospatial Data Production

Geospatial data production in the United States has traditionally been the purview of government agencies, which have been the only organizations with sufficient technical and financial resources to initiate complex, expensive, data collection and data production processes. Similarly, the United Kingdom's Ordnance Survey has been the center of national mapping in Great Britain, with a central role in collecting, producing, and licensing geospatial data. Geospatial data produced by government agencies in the US and the UK, and other government jurisdictions on a state and local level, is commonly described as *authoritative*, recognizing the central role of government organizations in generating such data.

Other sources of authoritative geospatial data include large map and atlas publishing firms such as Rand McNally, non-profit scientific and educational groups, such as the National Geographic Society, the United Nations, and large geospatial businesses such as GeoEye, TomTom, and Navteq. Each of these authoritative geospatial data producers, whether governmental, non-profit, or commercial, invests substantial resources in data production and quality control disseminating their data from a position of central authority.

This centralized authoritative production and distribution process contrasts sharply with the CGD processes introduced and defined in the previous chapter. The purpose of this chapter is to describe and contrast the authoritative, traditional geospatial data production methods and the CGD production methods, and to illustrate examples of hybrid approaches that use both production methods.

While *authoritative geospatial data* often has a clear lineage, production history, and in many cases, an error assessment (see Chapter 4 for a discussion of accuracy and error assessment), CGD is *asserted geospatial data* with few of the same characteristics. This distinction is important, as authoritative geospatial data may be perceived by users as being higher quality and more accurate data than asserted geospatial data, leading to a reluctance for Government agencies to accept asserted geospatial data or adopt crowdsourcing methods. Because of the collection of data to specifi-

cations and implementation of quality control checks, authoritative geospatial data may be assumed to always be error free, although a closer examination of this data reveals this is not the case.

Studies of positional accuracy (to be reviewed in Chapter 4) note that many CGD projects achieve accuracies comparable to authoritative sources. Goodchild notes that georegistration errors between authoritative sources and non-authoritative sources are often similar.[25] Even highly authoritative sources of geospatial data, contain positional and attribute errors and missing information, as noted in Figure 1 and Figure 2.

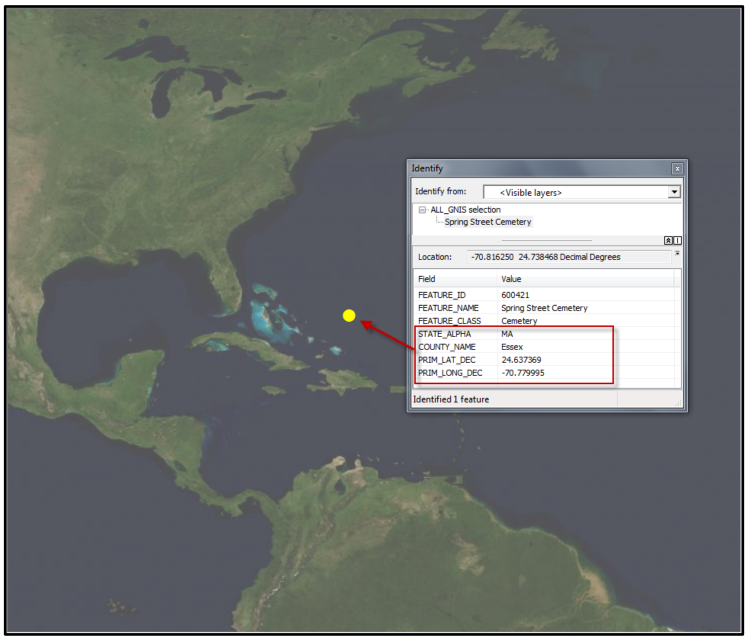


Figure 1. GNIS (USGS) graphic showing location of Spring Street Cemetery in Essex County, Massachusetts, incorrectly located in the Atlantic Ocean (April 2012)

[25] Goodchild, Michael F. "NeoGeography and the Nature of Geographic Expertise." *Journal of Location Based Services* 3, no. 2 (June 2009): 82–96.

GNS Extended View	
Name (Gazetteer Order/No Diacritics):	Sydney
DD Lat:	-33.861481
DD Long:	151.205475
Feature Designation Code:	PPLA
Feature Class:	P
Population:	NULL
Elevation (meters):	NULL
Creation Date:	1993/09/01
Modify Date:	2012/02/06
Effective Date:	NULL
Termination Date:	NULL

Figure 2. Information from NGA GEOnet Names Server, for Sydney, Australia missing information for population, elevation, effective date, and termination date (April 2012)

Goodchild[26] and others suggest that asserted geospatial data is often contributed by non-expert end-users for altruistic reasons, and that while noting the distinct differences with respect to lineage, quality assessment, and authority, asserted geosopatial data has significant benefits. Goodchild suggests that a primary benefit for asserted geospatial data is that it is produced by end-users with significant local expertise instead of by a central authority that may not be aware or have the capability of detecting changes in local environments.[27]

Goodchild and Glennon discuss the significant advantage that local geographic expertise poses during emergency events, such as the Santa Barbara wildfires of 2007-2009.[28] Zook et al. also underscore the benefits of crowdsourcing during the Haitian Earthquake of 2010, where a significant lack of geospatial data coverage for the impacted areas hampered initial rescue and support efforts.[29] Web-mapping services provided by end-users during the earthquake recovery were instrumental in supporting aid and relief agencies that were physically present in Haiti.

In addition to local geographic expertise and improved data coverage, asserted geospatial data can have better temporal coverage. During the Santa Barbara wildfires of 2008 and 2009, local citizens were able to contribute fire boundary updates in real-time through Google MapMaker, while the

26 Goodchild, "Citizens as Sensors: The World of Volunteered Geography."

27 Ibid.; Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content."

28 Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

29 Zook et al., "Volunteered Geographic Information and Crowdsourcing Disaster Relief."

local authoritative government mapping efforts had a much longer update cycle.[30]

Goodchild suggests that the authority of traditional mapping agencies can be attributed to their specifications, production mechanisms and programs for quality control.[31] Differences between authoritative geospatial data and asserted geospatial data are particularly evident in the techniques used for assessing and ensuring this quality and in the structure of the different communities associated with data production.

As noted, authoritative geospatial data is typically produced by governments, businesses, and organizations, with vast financial, technical, and organizational resources, while asserted geospatial data is produced by end-users, many of whom are untrained in typical geospatial fields. Goodchild suggests that the phenomenon associated with asserted geospatial data is related to a blurring of the roles between traditional, authoritative data producers and communicators, and the end-users referred to as neogeographers.[32]

The distinction between the authoritative elements of a scientific discipline and the layperson is usually very clear, and is, according to Goodchild, related to the complexity of the disciplines main concepts, the precise communication required by the discipline, and the high cost of making scientific observations. The financial, administrative, and educational requirements are so high, in areas such as particle physics, that the chance of any significant contribution by an untrained layperson would be remote. "No one would suggest that a neophysics might emerge that blurred the boundaries around high-energy physics; or that brain surgery might be invaded by a generation of untrained neoneurosurgeons."[33]

The emergence of neogeography and the phenomenon of CGD reflect a significant difference between traditional disciplines and the geospatial sciences. Goodchild states, "proximity to and familiarity with the subject matter of any science is a major factor in its public image and in the atti-

[30] Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

[31] Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content." p. 1.

[32] Ibid.; Sanjay Rana and Thierry Joliveau, "Neogeography Phenomena-Some Thoughts on It's Beginning, Future and Related Issues" (2007), http://www.ucl.ac.uk/~ucessan/ranajoliveauneogeogpapaper.pdf

[33] Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content." P. 2.

tudes that form around it . . . everyone feels himself or herself to be an expert in geography because geography is experienced by everyone.”[34]

Advancements in geotechnology, location aware mobile devices, cameras, mapping software, social media, and the Internet has generally increased interest in geospatial subjects and geospatial science[35] and importantly, has greatly reduced the cost of participation in geospatial science.

Goodchild cites many of these same factors, as well as the emergence of open-source geospatial software, as an important factor in leading to the emergence of neogeographers and production of asserted geospatial data.[36] While noting that the emergence of a community of largely untrained end-users actively producing asserted geospatial data has been perceived by the academic community and traditional data producers as a threat, Goodchild states that the academic community and traditional data producers have much to gain from the emerging neogeographic community, and the “activities, tools, and energies”[37] surrounding the emerging phenomenon.

What Goodchild sees as an emergent future, is one with a “potential for hybrid solutions, in which citizens and experts collaborate to combine their respective forms of expertise.”[38]

The Spectrum of Control in Geospatial Data Production

The following table provides a useful way to contrast extreme authoritative control over geospatial data production with a complete lack of control. At the extremes, anarchic systems produce lower quality information, while controlled systems produce higher quality information.

Anarchic systems encourage full and open participation with no guidelines or standards, no review and rapid release of data, while systems emphasizing control will limit the number of contributors, create products to predetermined specifications, incorporate a thorough review process, and control the release of data.

---

[34] Ibid.

[35] Rana and Joliveau, “Neogeography Phenomena-Some Thoughts on It’s Beginning, Future and Related Issues.”

[36] Goodchild, “Assertion and Authority: The Science of User-Generated Geographic Content.”

[37] Ibid.

[38] Ibid., p. 3.

Though possibly perceived as anarchic, many geospatial crowdsourcing projects have the elements of control typically associated with authoritative geospatial data collection. Crowdsourced efforts make conscious decisions where they fall on the anarchy-control continuum, balancing the natural tension between the quality of the data, the amount of control and the size of the crowd willing to participate.

Table 1. A Spectrum of control between Extreme Anarchy and Extreme Control

Extreme Anarchy	**Extreme Control**
No Contributor Expertise Required	Certified Technical Expertise Required
No User Registration	Verified User Registration
No Training Required	Certification Required
No Product Specification	Detailed Product Specification
No Production Practices	Established Production Practices
Any Geospatial Inputs	Approved Devices for Geospatial Input
No Specified Positional Accuracy/Precision	Specified Positional Accuracy/Precision
No Attribute Specification	Full Attribute Specification
Users Decide Which Features Collected	All Features Meeting Specification Collected
No Validation When Data Entered	Automated Point of Entry Validation
No Review	Professional Review
Multiple Users Edit A Feature	Single User Edits a Feature
No User Edit Tracking	Feature Level User Edit Tracking
No Edit Temporal Tracking	Edit Temporal Tracking
No Database Rollback	Database Rollback
No Metadata	Standards Compliant Metadata
Data Immediately Available	Data Available After Review
Unrestricted Data Availability	Proprietary Data
Unrestricted Usage Rights	Restricted Rights

OpenStreetMap – An Exemplar

OpenStreetMap (OSM)[39] (Figure 3) has the goal of creating a free worldwide map created by end-users. It is the most comprehensive asserted ge-

[39] "OpenStreetMap," n.d., http://www.openstreetmap.org/

ospatial data production project in existence, and is profiled in Chapter 3 of this report. With regard to the spectrum in Figure 3, OSM falls slightly left of the center of the spectrum for most parameters of geospatial data production and as a whole could not be characterized as extremely anarchic or extremely controlled. As a well-known and much-discussed CGD project, OSM has gained an element of authority because of its longevity and comprehensiveness.

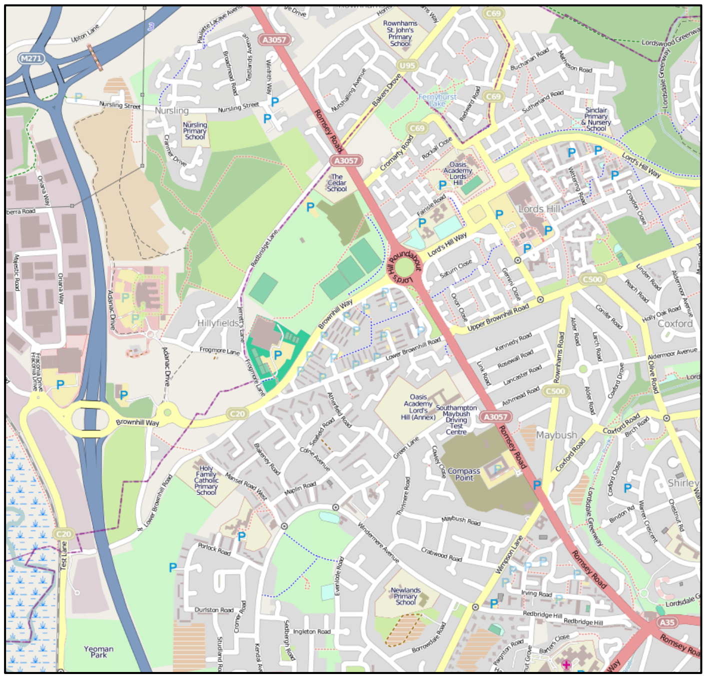


Figure 3. OSM coverage of Southampton, UK

As profiled by Goodchild[40] and others, OSM is a semi-organized, collaborative effort of volunteers, most of whom could be characterized as neogeographers, which is to say, have little formal academic training in geospatial fields but have an interest in geotechnology and open source projects. In OSM, many of the complexities of traditional map production are minimized or eliminated, and any complex fundamental issues are dealt with by the few highly-trained experts affiliated with the project. To explore OSM as an example of crowd-sourced geospatial data production, we use the facets referenced in Table 1 as a reference.

The majority of OSM contributors have no specialized technical expertise, though user registration is required to edit the data. No training is required, though a large body of wiki-based documentation exists and a us-

[40] Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content."

er-help center illustrates some relevant geospatial concepts and their implementation in OSM.

OSM has a large user community and as a result, there are recommended guidelines and practices for data production and editing, but there are no explicit or authoritative data specifications. Positional accuracy and precision are not explicitly specified or required, but are considered and adjusted as needed. Accuracy assessments for OSM data have been thoroughly reviewed by Haklay,[41] Girres et al.,[42] Mooney et al.,[43] Zielstra et al.,[44] and many others, and will be discussed in future chapters of this report. In the case of geospatial data precision, latitude and longitude values are often reduced to 6 or 7 decimal places (roughly equivalent to 10 centimeters of ground distance).[45] Some geospatial data attributes are defined through the user help documents. Users are permitted to enter any attribution, but OSM provides a set of recommended attributes and attribute values.

With regard to other notable elements of geospatial data production shown in Table 1, OSM allows users to decide which features are collected and makes data available immediately upon entry. It does have elements of review, where any user may edit another's work, but there is no professional review during the production process. Because of the project's notoriety, reviews of the OSM data production process and data quality have appeared in the peer-reviewed literature, but there is no internal peer-review process as occurs in highly controlled geospatial data production projects.

As of August 2012, OSM had unrestricted data availability, though its usage rights are governed by a Creative Commons license (CC-BY-SA) requesting attribution and share-alike provisions.[46] Commercial use of the

---

[41] M. Haklay, "How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets," *Environment and Planning. B, Planning & Design* 37, no. 4 (2010): 682.

[42] Jean-François Girres and Guillaume Touya, "Quality Assessment of the French OpenStreetMap Dataset," *Transactions in GIS* 14, no. 4 (August 2010): 435–459.

[43] P. Mooney, P. Corcoran, and A. Winstanley, "A Study of Data Representation of Natural Features in Openstreetmap," in *Proceedings of GIScience*, 2010, 150.

[44] D. Zielstra and A. Zipf, "Quantitative Studies on the Data Quality of OpenStreetMap in Germany," in *Proceedings GIScience*, 2010, 20–26.

[45] An example for dealing with precision in OSM can be found here:

"Script for Reducing the Precision of Nodes," Wiki, *OpenStreetMap Wiki*, May 15, 2010, http://wiki.openstreetmap.org/wiki/Script_for_reducing_the_precision_of_nodes

[46] "Attribution-ShareAlike 2.0 Generic (CC BY-SA 2.0)," *Creative Commons*, n.d., http://creativecommons.org/licenses/by-sa/2.0/

data is permitted under this license. OSM is in the process of moving to an Open Database License (ODbL). OSM is a Web-based, non-print mapping application, and incorporates elements of Web 2.0 data assemblage.

In the majority of geospatial data production areas, OSM tends to fall in the center of the spectrum between anarchy and control as elaborated in Table 1, or perhaps slightly to the left of center toward a project with less control. Notably, OSM tends toward the far left side in the areas of expertise and training, where none is required, and is far to the left in the lack of restrictions on data access and in the project's liberal data usage rights.

Because of its longevity, widespread use, and notoriety OSM has become an established entity within the crowdsourcing world, and has a few characteristics of authoritative geospatial data production, but by most measures and characteristics described in Table 1, OSM is more anarchic than authoritative, particularly when compared to geospatial data producers such as the Ordnance Survey of Great Britain (OS) or US Geological Survey (USGS). Both of these organizations would be characterized as having geospatial data production practices with extreme control.

Government Adoption of Crowdsourced Geospatial Data

After a period of initial skepticism, government agencies are now investigating ways to incorporate CGD under a variety of different models. These include 1) adopting non-Government crowdsourced data, 2) using CGD in parallel with authoritative data, and 3) integrated crowdsourcing methods and data. An important review of this topic can be found in Johnson et al. (2013).[47]

Under each of these models, the result could be characterized as a hybrid, where elements of authoritative and asserted geospatial data exist together. Goodchild suggests that "hybrid solutions to the production of geographic data may well represent the best of both worlds. There is clearly a role for central management and coordination, but the local expertise that VGI builds on is also very valuable."[48]

[47] Peter A. Johnson and Renee E. Sieber, "Situating the Adoption of VGI by Government," in *Crowdsourcing Geographic Knowledge*, ed. Daniel Sui, Sarah Elwood, and Michael Goodchild (Dordrecht: Springer Netherlands, 2013), 65–81, http://www.springerlink.com/index/10.1007/978-94-007-4587-2_5

[48] Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content." P. 16.

Adopting Non-Government Crowdsourced Data

An important emerging area for hybrid CGD projects, as discussed in the introduction to this report (Chapter 1), is in the area of emergency management, where response, recovery, and mitigation activities are often facilitated by volunteers and by CGD.

Goodchild and Glennon offer a research review motivated by California wildfire events,[49] while Zook et al. present a comprehensive review of the use of CGD and related information technologies in the aftermath of the devastating January 2010 Haitian earthquake.[50] Starbird,[51] and Starbird et al.[52] present informative perspectives on crowdsourcing dynamics during disaster response.

Zook et al. note that prior to the earthquake, large areas of Haiti lacked coverage by the Haitian government and by commercial geospatial data producers such as Google and Microsoft. The fundamental information needs that would typically be met by the government over the course of several years, i.e., detailed roadmaps, locations of critical assets, etc. were simply not available and economic conditions and had not presented a compelling reason for commercial firms to invest in detailed mapping of the country.[53]

Zook et al. dramatically underscore this issue by mapping the density of placemarks in Google Maps for the entire Island of Hispaniola for November 2009, just prior to the earthquake. Their maps shows a stark contrast, with a large number of placemarks in the Dominican Republic and very few on the Haitian side.[54] Because of the intense humanitarian interest, a

---

[49] Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

[50] Zook et al., "Volunteered Geographic Information and Crowdsourcing Disaster Relief."

[51] Kate Starbird, "What 'Crowdsourcing' Obscures: Exposing the Dynamics of Connected Crowd Work During Disaster," in *Collective Intelligence 2012* (presented at the Collective Intelligence 2012, Cambridge, MA, 2012), http://arxiv.org/abs/1204.3342

[52] Kate Starbird and Leysia Palen, "Pass It on?: Retweeting in Mass Emergency," in *Proceedings of the 7th International ISCRAM Conference* (presented at the ISCRAM, Seattle, WA: International Community on Information Systems for Crisis Response and Management, 2010), http://fsb.cvm.msu.edu/documents/starbirdpaleniscramretweet.pdf; Kate Starbird and Leysia Palen, "'Voluntweeters': Self-organizing by Digital Volunteers in Times of Crisis," in *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, CHI '11 (New York, NY, USA: ACM, 2011), 1071–1080, http://doi.acm.org/10.1145/1978942.1979102

[53] Zook et al., "Volunteered Geographic Information and Crowdsourcing Disaster Relief."

[54] Ibid.

social-media centered effort quickly emerged, whose goal was to build a geo-information infrastructure for Haiti.

The effort used CGD, and existing web-based mapping projects and services, such as CrisisCamp Haiti,[55] Ushahidi,[56] OSM, and GeoCommons.[57] OSM volunteers used open-source data, existing databases, and donated imagery to construct maps of buildings, transportation infrastructure, landmarks, and other features to provide volunteers on the ground with a geospatial framework to use in carrying out their essential activities.

Zook et al. cite the important role that the GeoCommons project had in providing data and information generated by both end users and by governments, and cite the project as an important hybrid with authoritative and asserted components. The US Department of Defense's Southern Command quickly adopted CGD in their coordinating role and provided a hybrid of authoritative and CGD through their All Partners Access Network (APAN),[58] which was an important central point for information dissemination and exchange.[59]

Using Crowdsourced Geospatial Data in Parallel with Authoritative Data

Another notable example of a hybrid geospatial data production project is the National Geospatial-Intelligence Agency's (NGA) PLACES program, which is an effort to capture vernacular place name references through crowdsourcing.

The PLACES data will be stored, accessed, and visualized separately from authoritative data in the GEOnet Names Server in order to prevent any confusion about the source of the names.

In this approach, the raw crowdsourced data is neither directly adopted as official nor integrated with authoritative data, but is accessible to users in

[55] "Connecting People, Tools and Resources to Support Crisis Response," *CrisisCommons*, 2012, http://crisiscommons.org/; "CrisisCamp Haiti - Washington DC," *Eventbrite*, 2012, http://crisiscamphaitiwdc.eventbrite.com/.

[56] "Ushahidi," *Ushahidi*, n.d., http://ushahidi.com/

[57] "GeoCommons," *Geocommons*, n.d., http://geocommons.com/

[58] "APAN Community," n.d., https://community.apan.org/default.aspx

[59] Ibid.

parallel with authoritative data. Crowdsourced data entries may, however, be evaluated and incorporated in the authoritative database after review by professional toponymists.

The PLACES program and other similar research efforts such as Rice et al. (2012) and Twaroch et al. (2009) use crowdsourcing for building hybrid systems of authoritative and asserted placenaming. Kostanski's 2011 and 2012 reports on crowdsourcing applied to gazetteers are notable,[60] reflecting some of the same approaches suggested by Rice et al.[61]

Integrating Crowdsourcing Methods and Data

The National Oceanographic and Atmospheric Administration's (NOAA) National Weather Service SKYWARN program[62] is an effort to gather severe weather reports from network of nearly 300,000 trained severe weather spotters, who provide information about local storms, flooding, and other weather conditions. The crowdsourced reports from these weather spotters are used in a hybrid approach to refine, update, and validate weather forecasts, warnings, and alerts issued by the National Weather Service.

SKYWARN volunteers are recruited from the ranks of fire fighters, emergency medical service technicians, dispatchers, utility workers, and local citizens, and trained (at no cost to the volunteer) at their local weather forecast offices. The two hours of training includes the basics of storm development and identification, safety procedures, and reporting protocols. A notable aspect of the SKYWARN program is its longevity, having started in the 1970s.

Over the last 20 years, the USGS has initiated several hybrid geospatial data production projects, some of which continue from much earlier efforts

---

[60] Laura Kostanski, *To Study the Methods for Recording Unofficial Place Names into Comprehensive Data Sets for Improvement Knowledge Transfer*, Technical Report (Australia: The Winston Churchill Memorial Trust of Australia, 2011), http://www.churchilltrust.com.au/site_media/fellows/2011_Kostanski_Laura.pdf; Secretary Committee for Geographical Names of Australia, Australia and Laura Kostanski, "Crowd-Sourcing Geospatial Information for Government Gazetteer," in *Tenth United Nations Conference on the Standardization of Geographical Names* (presented at the Tenth United Nations Conference on the Standardization of Geographical Names, New York, NY, 2012), http://unstats.un.org/unsd/geoinfo/ungegn/docs/10th-uncsgn-docs/crp/E_Conf.101_CRP16_Summary%20paper%20of%20VGI%20for%20UNGEGN.pdf

[61] Rice et al., "Supporting Accessibility for Blind and Vision-impaired People With a Localized Gazetteer and Open Source Geotechnology."

[62] "What Is SKYWARN?," *NWS SKYWARN*, May 10, 2011, http://www.nws.noaa.gov/skywarn/

to get feedback and cartographic updates from end-users, and others that involve newer Web 2.0 techniques.

The first project was The National Map Corps, which was initiated back in 1994 under the Earth Science Corps name, and involved end-users adopting a specific 7.5 minutes USGS topographic quadrangle map and adding annotations, corrections, and updates.[63] This project was renamed the National Map Corps in 2001 during the rollout of the USGS's National Map, and involved more than a thousand volunteer members collecting and contributing tens of thousands of updates and corrections via spreadsheets. In 2006 the project transitioned to a web-based workflow involving hundreds of volunteers, but in 2008 the project was suspended due to funding limitations.

A more recent version of hybrid geospatial data production by the USGS continues the National Map Corps name and mission, but adopts the Web 2.0 collaborative framework of OSM for the generation of data using volunteers, but not the OSM data. The OSM Collaborative Prototype (OSM CP) uses the OSM online editor with USGS data, with updates sent to the National Map.[64] This new effort is still in its formative stages and an early report on the project suggests that OSM software is an effective way for USGS to do collaborative editing and incorporate CGD into the National Map.[65]

The next chapter of this report takes an in-depth look at a number of CGD projects and applications. The projects and applications profiled in Chapter 3 are not intended to be a comprehensive census of the domain, but instead have been selected by the authors and their collaborators to represent a wide range of applications, and a broad spectrum of activities. The applications profiled will help the reader develop a sense of the significant developments happening in the area of CGD.

---

[63] See "The National Map Corps," *USGS*, August 2, 2011, http://nationalmap.gov/TheNationalMapCorps/; "History of Volunteer Mapping at the USGS," *USGS*, August 2, 2011, http://nationalmap.gov/TheNationalMapCorps/history.html

[64] "This Is the Home of The National Map Corps," *USGS*, May 9, 2012, https://my.usgs.gov/confluence/display/nationalmapcorps/Home

[65] Eric B. Wolf et al., *OpenStreetMap Collaborative Prototype, Phase One*, Open--File Report (U.S. Geological Survey, Reston, VA: U.S. Department of the Interior, U.S. Geological Survey, 2011), http://pubs.usgs.gov/of/2011/1136/pdf/OF11-1136.pdf

3 Crowdsourced Geospatial Data Case Studies

Introduction

The development of social media and Web 2.0 over the last decade has led to many changes in the way information is created and shared, with an emphasis on participation, sharing, and collaboration. Although these developments are relatively recent, a large number of crowdsourced geospatial projects have emerged during this period.

Some projects presented in this chapter are relatively new, while others can trace origins back several decades, as noted in the previous chapter's discussion of the National Weather Service SKYWARN program and the USGS's National Map Corps.

The projects discussed in this chapter are a sampling from the hundreds of geospatial crowdsourcing applications and range from small efforts involving tens of contributors to massive communities with millions of members. They cover the range of activities, from the acquisition of raw imagery over small areas, to building a worldwide, openly available database. Some projects, like OSM and Google MapMaker produce geospatial framework data similar to that generated by national mapping agencies. Geospatial crowdsourcing is not limited to framework data, however, and may be applied to any content that can be geolocated, including short text messages (Twitter), photographs (Flickr) and encyclopedia entries (Wikipedia).

Often, a single project supports multiple geospatial data collection tasks. For example, Grassroots Mapping provides guidance and equipment for kite and balloon imagery acquisition, as well as tools to georeference the resulting imagery. Another example, OSM, perhaps the largest and most widely known CGD project, supports digitizing, attributing, and validating functionality.

Most geospatial data collection projects involve crowd creation; but there are also examples of crowd voting (SurveyMapper), crowd wisdom (Defense Advanced Projects Agency (DARPA) Network Challenge), and crowd funding (Balloon Mapping Kit from Grassroots Mapping). While the ma-

jority of applications are software driven, one open hardware effort (Balloon Mapping Kit from Grassroots Mapping) is discussed.

Geospatial data collection can be dissected into a number of tasks or steps, almost all of which have been crowdsourced. Table 2 classifies geospatial data collection tasks, describes them, and highlights sample projects or applications related to the task. The following examples, themes and ideas addressed here will highlight the potential of geospatial crowdsourcing and will carry through to the following chapters on spatial data quality, evaluating CGD, and lessons learned. This chapter provides an overview of these projects, while detailed descriptions are available in Appendix 2.

Table 2. Geospatial crowdsourcing applications

Tasks	Description	Example
Imaging	Building collections of imagery.	• Grassroots Mapping
Georeferencing	Rectifying maps and imagery.	• Grassroots Mapping • NYPL Map Rectifier
Transcribing	Converting text resources to a digital form.	• OldWeather
Digitizing	Collecting geospatial feature geometry and attributes from maps or imagery.	• OSM • Google MapMaker • Wikimapia
Attributing	Adding descriptive information to known geospatial features or datasets.	• Galaxy Zoo
Reporting	Collecting information about a location, usually through observation or a mobile device.	• Louisiana Bucket Brigade • GasBuddy • Street Bump • SyriaTracker • Wikipedia
Searching	Searching maps or imagery to identify specific features.	• Field Expedition: Mongolia - Valley of the Khans Project • DARPA Red Balloon
Tracking	Collecting paths and traces, usually using GPS.	• Waze
Validating	Verifying the quality of existing geospatial information.	• NAVTEQ Map Reporter • Geo-Wiki.org • OSM Inspector
Polling/Surveying	Collecting place-based opinions or information from users.	• SurveyMapper
Socializing	Contributing geospatially referenced in-	• Twitter

	formation to social media sites.	• Flickr • Foursquare
Sharing	Placing content on hosted site, potentially including data, applications, or finished maps, where users can access and mash-up.	• ArcGIS Online • GeoCommons

Imaging

Building an imagery collection has traditionally been an expensive and resource intensive application, requiring airborne collection platforms, processing facilities, large amounts of storage, and data dissemination resources. The general public rarely saw or interacted with this imagery, due to the high cost and requirement for specialized software. In 2005, Google changed public interaction with imagery collection. The release of Google Maps delivered satellite imagery to all web browsers. The widespread availability of imagery and associated maps revolutionized geography, bringing resources to the general public that were previously available only to select users (i.e. academic, military, government, etc.).

Image collection via crowdsourcing is now possible at local scales due to the emergence of hyperlocal image collection, where high resolution, current imagery is collected using low cost, simple platforms. Despite these advances, however, attempts to crowdsource extensive image collections, like Open Aerial Map,[66] have failed. Acquiring and disseminating large collections remains a task for large companies and Government agencies.

Grassroots Mapping[67]

Grassroots Mapping combined crowd funding with open hardware to create a balloon mapping platform for aerial imagery collection. Raw imagery collected from the balloon mapping platform can be georegistered and made available in the public domain through Public Laboratory's open data archive: PLOTS.[68]

Grassroots Mapping was developed for the Deepwater Horizon Oil Spill in 2010 by building on the traditions of kite and balloon mapping. The goal was to empower local citizens in the cleanup effort, with the hopes that the

[66] "OpenAerialMap," March 23, 2011, http://openaerialmap.org/Main_Page

[67] "Grassroots Mapping," March 9, 2012, http://grassrootsmapping.org/

[68] "The PLOTS Archive," *Publiclaboratory.org*, April 22, 2012, http://publiclaboratory.org/archive

data would also be valuable to scientists. Contributors captured imagery over Louisiana, Mississippi, Alabama and Florida, using Balloon Mapping Kits purchased online.

Figure 4. Balloon mapping kit in action.[69]

The Balloon Mapping Kit (Figure 4) is just one component of the Grassroots Mapping solution. Once imagery is collected, it needs to be georeferenced so that it can be fused with other geospatial data

Georeferencing

Aligning data sources to known geographic or projected coordinate systems is fundamental to mapping and geospatial analysis. It provides the link between the pixels in an image or scanned map and the real world. Once data are georeferenced, they can be overlaid with other geographic data. Due to the technical complexity of georeferencing, these applications draw fewer contributors.

Two crowdsourced georeferencing projects stand out: the previously-highlighted Grassroots Mapping for registering imagery and the New York Public Library Map Rectifier for georegistering maps.

Grassroots Mapping[70]

Georeferencing imagery collected from kites or balloons can be accomplished using the web-based MapKnitter application,[71] which is a free and open source software application. MapKnitter is not a full-featured orthorectification system that removes camera and terrain distortions for high positional accuracy. It does, however, work well as a lower accuracy registration capability for kite and balloon imagery. The results are suitable for overlay and visualization in applications like Google Maps (Figure

[69] Louisiana Bucket Brigade, *Untitled*, from Flickr.com, JPEG Image, 540 x 720 pixels, January 1, 1980, http://farm6.staticflickr.com/5041/5244604427_683927c894.jpg

[70] "Grassroots Mapping."

[71] "PLOTS Map Knitter," n.d., http://mapknitter.org/

5), where the imagery resolution is typically lower than the user-collected imagery.

Figure 5. Balloon imagery[72] overlayed on Google imagery

New York Public Library (NYPL) Map Warper[73]

The NYPL Map Warper application allows contributors to rectify historical maps from the NYPL collections to match current maps; and then makes them available to the public (Figure 6).

[72] *39646.png*, from Publiclibrary.org, PNG Image, 256 × 256 pixels, December 7, 2011, http://archive.publiclaboratory.org/wcu/2011-12-07-northcarolina-cullowhee-westerncarolinauniversity/tms/16/17625/39646.png

[73] "NYPL Map Warper: Home," n.d., http://maps.nypl.org/warper/

Figure 6. Historical map of Manhattan[74] overlaid on Google Imagery. This image shows a 1915 redrafting of the 1660 Castello Plan of lower Manhattan overlaid on current imagery. The expansion of the island is clearly shown in this illustration[75]

Anyone can contribute to the site once registered and they can register new maps or improve the registration of existing maps. To register a map, a contributor identifies control points, which are common locations on the historic and modern map. The control points are input to the rectification software which warps the historical map to the current map. Map Warper automatically calculates errors for control points and if the error is significant, displays the control point in red, so the user can correct the error.

Transcribing

Transcribing enables contributors to copy information from documents and record it in a digital form. These projects rely primarily on the labor of

[74] "NYPL Map Warper: Viewing Map 13913," n.d., http://maps.nypl.org/warper/mapscans/13913

[75] Matt Knutzen and Stephen A. Schwarzman, "Drawing on the Past: Enlivening the Study of Historical Geography at Maps.nypl.org," Blog, *NYPL Labs*, February 3, 2010, http://www.nypl.org/blog/2010/02/03/drawing-past-enlivening-study-historical-geography-mapsnyplorg

the crowd, not its wisdom. An example of this type of effort is the Transcribe Bentham project[76] of the University College of London, which employed the crowd in copying the 12,400 manuscripts of Jeremy Benthem, a British utilitarian philosopher.

Transcription has also been used to document natural history collections in projects like Notes for Nature, where historical ledger pages, annotated images, and specimen have been converted to digital form. Transcription tasks often focus on entire documents or selected information from the documents.

Old Weather[77]

Old Weather is an innovative geospatial crowdsourcing transcription project (Figure 7). It is a model example for engaging users, ensuring quality data, and applying the data to scientific problems. Contributors transcribe information from World War I era Royal Navy ship logbooks in order to fill data gaps in climate change research.

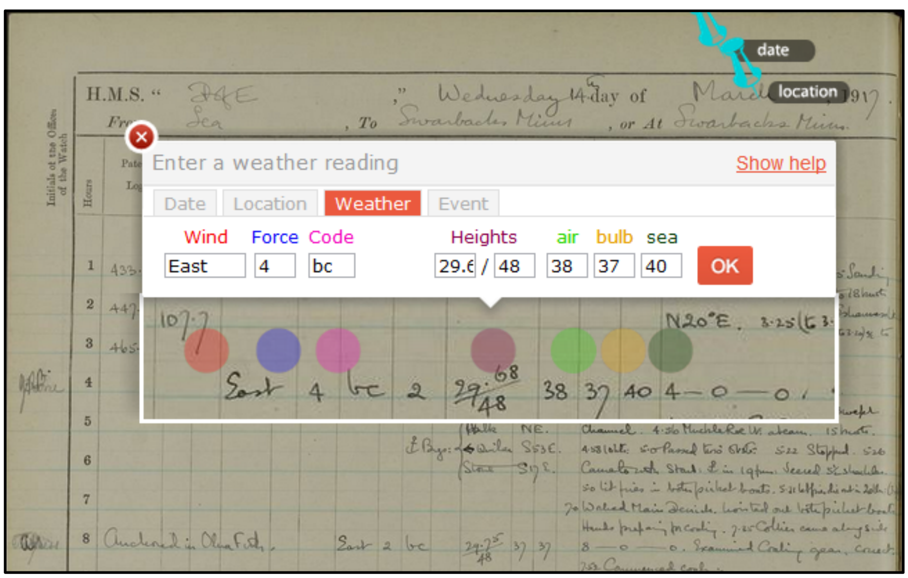


Figure 7. Sample Logbook Data Entry. Old Weather guides help users enter the appropriate data. This weather guide shows colored circles near the appropriate locations on the page, aiding the transcribing process and reducing errors (Image by author)

[76] "Transcribe Bentham," n.d., http://www.ucl.ac.uk/transcribe-bentham/

[77] "Old Weather - Our Weather's Past, the Climate's Future," n.d., http://www.oldweather.org/

Old Weather has made significant progress since its start. As of August 2012, it had over 27,000 contributors capturing more than a 1.6 million weather observations data from ships logs. Based on their estimates, it would take 28 years for one individual to extract all the information from a single log. Therefore crowdsourcing allows the researchers to meet their requirements in considerably less time.

Old Weather incorporates a number of different mechanisms in order to ensure quality data. Researchers compare transcriptions of the same logbook entry by a minimum of three different contributors in order to verify the data values. Even when an entry is transcribed correctly, the data may contain errors. Therefore, automated checks are made using valid data ranges.

Digitizing

Digitization is a traditional method for collecting geospatial data that many national mapping agencies use, when analysts collect feature geometry, attributes, and topology from maps and imagery. Although automated feature extraction has evolved significantly, manual digitizing remains the preferred method for collecting geospatial data that requires interpretation. Due to its labor-intensive nature, this task is ideal for crowdsourcing.

Three ambitious, global digitizing efforts are reviewed below: OSM, Google MapMaker, and Wikimapia. Particular attention is paid to production and distribution processes, as these highlight the variety of possible approaches to crowdsourcing data. While all have the goal of digitizing the world, each application approaches this task in different ways; specifically in the data review processes, release of contributed data, and licensing (refer to Table 3).

Table 3. Comparison of OSM, Google Map Maker, and Wikimapia

Project	Review	Feature Locking	Release of Data	Distribution	Licensing
OSM	Users	No	Instantaneous	Map tiles, download, and API	Creative Commons Attribution-ShareAlike license transitioning to Open Database License
Google Map Maker	Users, hierarchy of editing privileges based on experience, and Google staff	Yes	Varies, edits may be delayed for review	Map tiles	Proprietary to Google
Wikimapia	Users, hierarchy of editing privileges based on experience, and Wikimapia Administrators	Limited	No information	API	Creative Commons Attribution-NonCommercial-ShareAlike

OpenStreetMap[78]

A groundbreaking crowdsourced application, OSM, initiated the crowdsourcing paradigm in the geospatial community. The purpose of the OSM project is to create an open access map of the world that could be edited by anyone (Figure 8).

OSM originated in the United Kingdom in 2004 as an alternative to Ordnance Survey data, which was covered under 'Crown Copyright' and had expensive licensing fees that severely restricted its use.[79] Data licensing and availability in the United Kingdom were very different than in the United States, where most Federal government data was available at no cost and with no licensing restrictions.

Initially, volunteers used Global Positioning System (GPS) technology to collect street centerline data and produce maps that were freely available, and allowed anyone to make modifications or updates to the existing information. This was later supplemented with an aerial imagery base, provided by Yahoo and later by Microsoft Bing.

[78] "OpenStreetMap."

[79] "Ordnance Survey - OpenStreetMap Wiki," n.d., http://wiki.openstreetmap.org/wiki/Ordnance_Survey

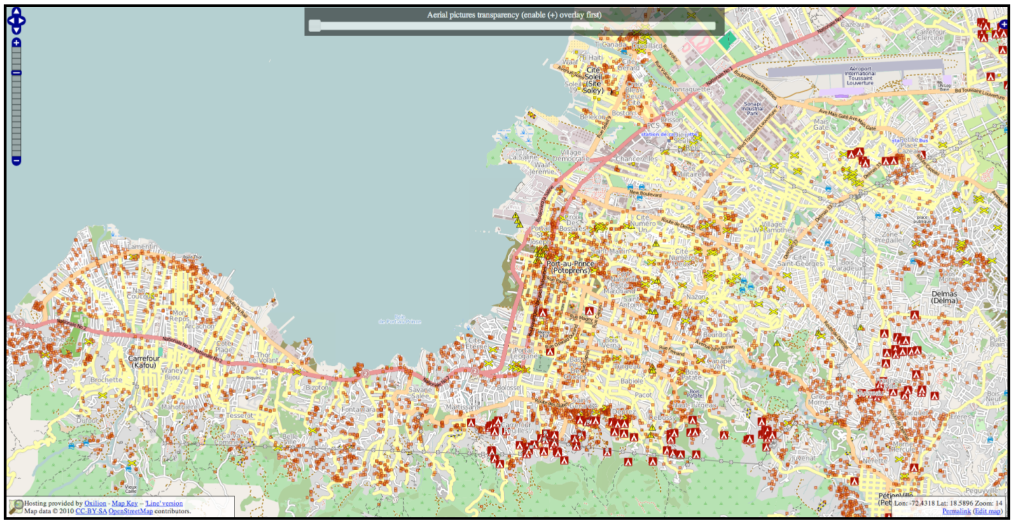


Figure 8. This is a screenshot taken from the OSM of Port-au-Prince Haiti[80]

OSM has a rich environment for collecting, editing, and verifying nodes (points), ways (lines and polygons), and relations (ordered lists of nodes, ways, and other relations). The amount of data collected is impressive. As of August 2012, OSM had over 30 billion GPS points and 14.5 billion ways.[81]

OSM data is available to the public as soon as it is entered. There is no formal review structure beyond the edits and reviews done by other contributors. As noted in Chapter 2, as of August 2012 it was licensed under the Creative Commons license (CC-BY-SA) and will be moving to an ODbL.

OSM is perhaps the gold standard case of geospatial crowdsourcing, and serves as a great example for similar efforts. It is the most successful effort, having been adopted by major companies like Apple, MapQuest, and Foursquare. Yet it is also very unique as the only geospatial crowdsourced effort to achieve this level of success, where it is considered and has been adopted as an alternative to traditional mapping data.

80 "Haiti Crisis Map - OpenStreetMap NL," n.d., http://haiti.openstreetmap.nl/

81 "OpenStreetMap Statistics."

Google Map Maker[82]

Google Map Maker (Figure 9) is a Google service intended to utilize the crowd to expand the geographic content of Google Maps and Google Earth, as well as other Google Products, such as Google Places.

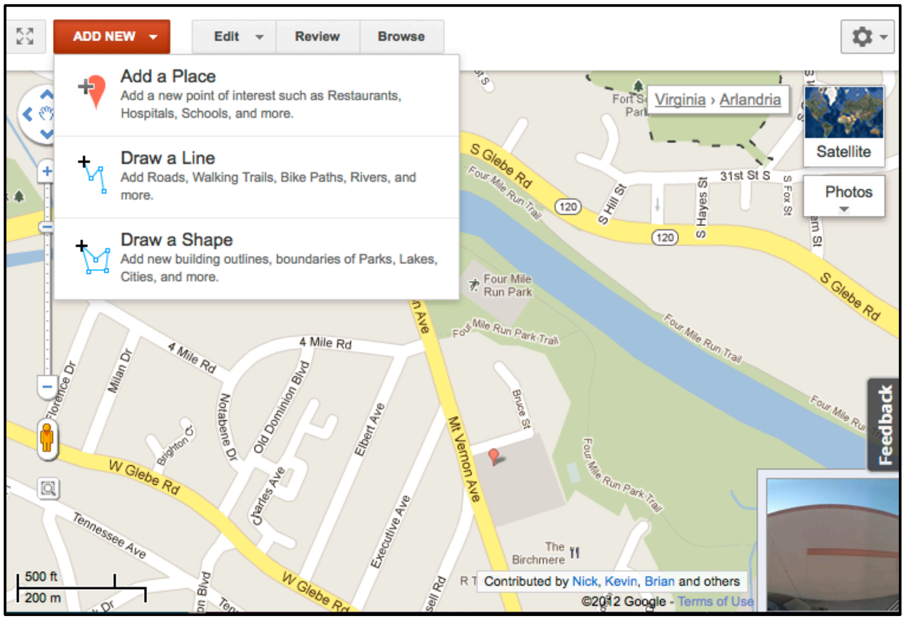


Figure 9. Google Map Maker user interface[83]

Google uses a system of contributor review, but retains final approval authority for the content. The Google team exercises the right to lock features to prevent user editing. For example, transit features in the United States are locked.[84]

Any user can view the map tiles displaying information created in Google Map Maker through Google Maps or Google Earth. However the underlying data is proprietary in the sense that it cannot be downloaded or accessed through an Application Programming Interface (API). When users contribute data, they give Google “a perpetual, irrevocable, worldwide, royalty-free, and non-exclusive license to reproduce, adapt, modify, trans-

82 “Google Map Maker,” *Google*, n.d., http://www.google.com/mapmaker

83 Ibid.

84 “‘This Feature Has Been Locked’ - Google Groups,” n.d., https://groups.google.com/forum/?fromgroups#!topic/google-mapmaker/tykoUwykD3Q

late, publish, publicly perform, publicly display, distribute, and create derivative works of the User Submission."[85]

Wikimapia[86]

Wikimapia's motto "Let's describe the whole world!" underlies their goal to "create and maintain a free, complete, multilingual, up-to-date map of the earth's surface."[87] The site contained over 18.4 million places and 1.6 million registered users in May 2012.[88] Wikimapia differs from other general mapping efforts by focusing on places, a broad notion covering everything from buildings to parks and communities (Figure 10).

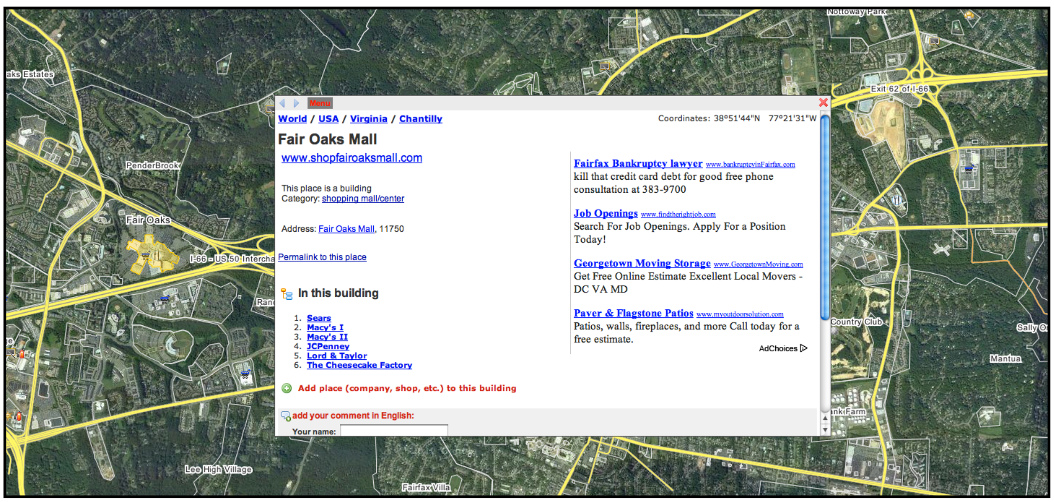


Figure 10. Screenshot taken from Wikimapia highlighting some of its functionality. The information seen in this screenshot is for Fair Oaks Mall in Fairfax, VA[89]

Goodchild and Glennon noted that Wikimapia has been subject to repeated and significant malicious content, leading to a decline in its reputation.[90] Due to these significant issues with malicious and mischievous contributions, Wikimapia has extensive procedures related to the treatment of vandalism and the banning of users.

---

[85] "Google Map Maker Terms of Service," n.d., http://www.google.com/mapmaker/mapfiles/s/terms_mapmaker.html

[86] "Wikimapia - Let's Describe the Whole World!"

[87] "User Guide: Philosophy - Wikimapia," n.d., http://wikimapia.org/wiki/User_Guide:_Philosophy

[88] "Wikimapia - Let's Describe the Whole World!"

[89] Ibid.

[90] Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

While contributors give Wikimapia unrestricted rights to use their contributions, the data is available at no cost through an API. Wikimapia licenses their data with a Creative Commons Attribution-NonCommercial-ShareAlike license, which prohibits commercial use. This is a significant difference from the licensing employed by OSM, which allows for commercial use. In addition, there remain issues with using Wikimapia data, as it was digitized over Google imagery and could potentially fall under derived-content copyright restrictions.

Attributing

Attributing tasks offer tools to describe the characteristics of geospatial features whose location is already known. Attribution tasks are simpler and more focused than digitizing tasks, which include the collection of feature geometry and attributes.

Galaxy Zoo[91]

Galaxy Zoo is an excellent example of a site focusing on the attribution of features. Contributors to Galaxy Zoo analyze the shapes of galaxies identified in Hubble telescope imagery (Figure 11). The task is well suited to human interpreters, who classify the galaxies by answering a series of simple questions. Results of the effort have impacted the study of space, leading to the redirection and refocusing of earth and spaceborne telescopes on galaxies of interest.

Galaxy Zoo demonstrates the power of geospatial crowdsourcing for collecting substantial amounts of data in a very short time.

> The original Galaxy Zoo was launched in July 2007, with a data set made up of a million galaxies imaged with the robotic telescope of the Sloan Digital Sky Survey. With so many galaxies, the team thought that it might take at least two years for visitors to the site to work through them all. Within 24 hours of launch, the site was receiving 70,000 classifications an hour, and more than 50 million classifications were received

[91] "Galaxy Zoo: Hubble," n.d., http://www.galaxyzoo.org/

> by the project during its first year, from almost 150,000 people.[92]

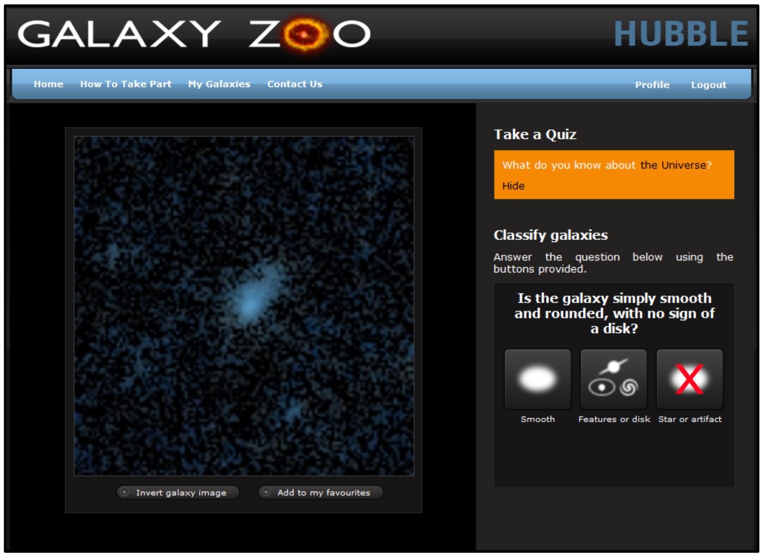


Figure 11. Galaxy Zoo's attribution screen. The contributor merely looks at an image and selects the appropriate answer by clicking on a button.[93]

Quality control is based on multiple assessments of the same galaxy images. Galaxy Zoo controls the dissemination of images to the contributors.

> Having multiple classifications of the same object is important, as it allows us to assess how reliable each one is. For some projects, we may only need a few thousand galaxies but want to be sure they're all spirals. No problem - just use those that 100% of classifiers agree on. For other projects we might want larger numbers of galaxies, so might use those that a majority say are spiral.[94]

Galaxy Zoo is a model example of a site dedicated to attribution, keeping the task simple while providing an interesting and engaging environment for large numbers of contributors.

---

[92] Ibid.

[93] "Galaxy Zoo: Classify Galaxies," n.d., http://www.galaxyzoo.org/#/classify

[94] "Galaxy Zoo: Hubble - The Story So Far," n.d., http://www.galaxyzoo.org/#/story

Reporting

In his landmark 2007 paper titled 'Citizens as sensors: the world of volunteered geography,' Goodchild outlined a vision where average citizens become sensors, measuring and observing the world to create a global geographical understanding.[95] With the recent widespread availability of smartphones, devices that can capture images, video, audio, time, location, and other observations, such as motion, this vision is rapidly becoming a reality, supplementing and surpassing traditional reporting.

Reporting may be one of the most viable crowdsourced applications, as it harnesses the local knowledge and observations of individuals who widely distributed across space and time, individuals with ready access to devices that can collect and disseminate data.

Reporting applications have been successfully implemented across a broad range of applications, from reporting natural disasters to monitoring elections to identifying potholes. Examples of geospatial crowdsourced reporting highlighted in this section include collecting environmental data with Louisiana Bucket Brigade,[96] sharing local gas prices with GasBuddy,[97] automatically identifying potholes with Street Bump,[98] and reporting crime with SyriaTracker.[99]

Louisiana Bucket Brigade[100]

The Louisiana Bucket Brigade is an environmental health and justice organization that works with local citizens to monitor air quality. Data is collected using 'buckets,' which are low cost, easy-to-use, air-sampling devices that are government approved (Figure 12). The goal is to empower fence-line neighbors, who border industrial facilities, to collect scientifically valid samples that are recognized by agencies that regulate industrial pollution.

---

[95] Goodchild, "Citizens as Sensors: The World of Volunteered Geography."

[96] "LA Bucket Brigade : Index," n.d., http://www.labucketbrigade.org/

[97] "GasBuddy.com - Find Low Gas Prices in the USA and Canada," n.d., http://gasbuddy.com/

[98] "Street Bump," n.d., http://streetbump.org/.

[99] "Syria Tracker," *Syria Tracker: Missing, Killed, Arrested, Eyewitness, Report*, n.d., https://syriatracker.crowdmap.com/.

[100] "LA Bucket Brigade : Index."

Figure 12. Taking an air sample[101]

The bucket brigade process illustrates the importance of organization in crowdsourcing efforts, requiring coordination among volunteers who take on different roles: sniffers identify problems, samplers take air measurements, and coordinators collect and replace samples.[102]

The Louisiana Bucket Brigade has achieved some success, notably the identification of high levels of chemicals in Norco, LA. Volunteer samples detected levels of methyl ethyl ketone (MEK) and benzene that violated Louisiana state standards.[103]

GasBuddy[104]

GasBuddy is a gas price reporting application that enables individuals to contribute and search for information pertaining to gas prices at local gas stations. Given the volatility of gas prices and locally varying costs people have a keen interest in having accurate and timely data regarding the location for the cheapest gas.

---

101 Louisiana Bucket Brigade, *Taking an Air Sample*, from Facebook.com, JPEG Image, 558 × 371 pixels, February 2005, http://sphotos.xx.fbcdn.net/hphotos-ash4/292196_10150641900818963_289341778962_9317550_1210779005_n.jpg.

102 Dara O'Rourke and Gregg P. Macey, "Community Environmental Policing: Assessing New Strategies of Public Participation in Environmental Regulation," *Journal of Policy Analysis and Management* 22, no. 3 (2003): 383–414 (389).

103 "LA Bucket Brigade : Air Sample in Norco's Diamond Neighborhood Shows Extreme Levels of Chemicals," n.d., http://www.labucketbrigade.org/article.php?id=803.

104 "GasBuddy.com - Find Low Gas Prices in the USA and Canada."

Gas Buddy is both space and time sensitive. As a means to keep the information relevant and accurate, GasBuddy has a policy that requires them to remove all prices that exceed a 72-hour time frame. Figure 13 provides a screenshot of the reporting display of the GasBuddy application while Figure 14 shows the map display.

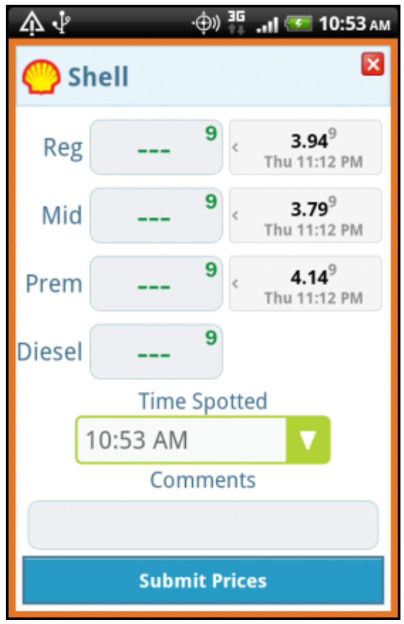


Figure 13. Mobile screen for gas prices report in GasBuddy[105]

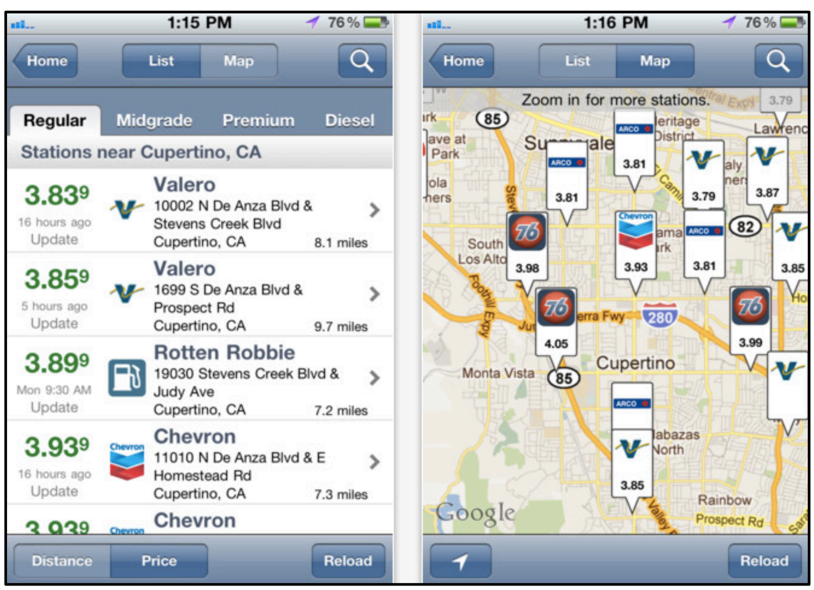


Figure 14. GasBuddy locations: list and map display[106]

105 "GasBuddy: Save Dollars at the Pump," *Android.AppStorm*, n.d., http://android.appstorm.net/reviews/lifestyle/gasbuddy-save-dollars-at-the-pump/.

Street Bump[107]

Street Bump is a free mobile phone application for collecting road smoothness information in the city of Boston. The application records information about the bumpiness of the ride that can be used to identify potholes.

Street Bump makes creative use of smartphone sensors. Bumps in the road are detected using the smartphone's accelerometer and located using the integrated GPS system. Initial experiments identified a difficulty differentiating between manholes, potholes, and other bumps in the road.

Street Bump relies on machine-to-machine communication to passively report information. Unlike other pothole applications, such as the City of Toronto's online reporting[108] or their certified smartphone application,[109] Street Bump automates the process, relieving the contributor from having to type text, add photographs, or operate the phone while detecting potholes. Contributors simply turn on the application and it runs automatically, removing any risk of distracted driving.

The Street Bump concept, if not the actual application, represents one future for crowdsourced, location-based sensing by capitalizing on the sensors in the smartphone and passively collecting and transferring data.

SyriaTracker[110]

SyriaTracker is a citizen crime-reporting site focusing on violence in Syria, with reports covering missing persons, killings, arrests, and other crimes. SyriaTracker is unique in the flexibility of user input, allowing contributors to report crimes in multiple languages and multiple channels, including: direct web entry (See Figure 15), sending reports through a smartphone,

---

106 "GasBuddy - Find Cheap Gas Prices," *iTunes Store*, October 14, 2011, http://itunes.apple.com/us/app/gasbuddy-find-cheap-gas-prices/id406719683?mt=8.

107 "Street Bump."

108 "Self-service," *311 Toronto*, April 25, 2012, https://secure.toronto.ca/webwizard/html/pothole_repair.htm.

109 "TDOT 311," *Public Leaf*, n.d., http://www.publicleaf.com/tdot311.

110 "Syria Tracker."

sending reports via email, tagging Twitter tweets with hashtags, or using the Google Speak2Tweet[111] service to call a phone number and leave a message.

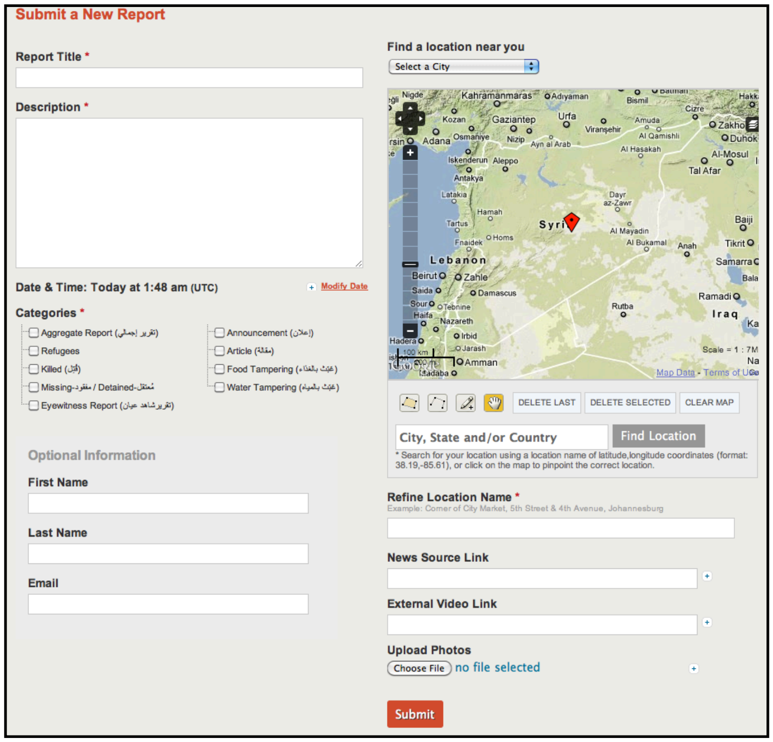


Figure 15. SyriaTracker Web-based report form[112]

Syria Tracker is built on the Ushahidi platform[113] an open source application designed specifically for crowdsourcing. The company's software has been used to document a wide range of events, from Snowmaggedon in New York City to monitoring voting in India to documenting survivor needs for the Japanese tsunami.

Wikipedia[114]

Wikipedia is a free, online, multilingual encyclopedia that allows users to find, edit and publish information. By August 2012, Wikipedia had over 17 million contributors with more than 4 million content pages[115] and was ranked by Alexa (a company that provides services and tools for dynamic

---

111 "Speak To Tweet (speak2tweet) on Twitter," n.d., https://twitter.com/speak2tweet.

112 "Submit a New Report," *Syria Tracker*, n.d., https://syriatracker.crowdmap.com/reports/submit.

113 "Ushahidi."

114 "Wikipedia," n.d., http://www.wikipedia.org/.

115 "Statistics."

Web navigation) as the sixth-most frequently accessed web site in both the United States and the world.[116] Geography-related topics consistently rank high in Wikipedia. According to a 2007 study, geography was the third most popular topic and represented 12% of the most frequently visited Wikipedia pages.[117] Geographic articles are organized topically,[118] and articles about specific places or features are often referenced by lists (such as the list of "cities in Afghanistan"[119] or the list of "mountains"[120]).

Individual articles on geographic topics contain free-form text as well as structured infoboxes or fixed format tables of information. There are a number of different types of infoboxes depending on the content of the article. The infobox shown in Figure 16 is tailored for articles about mountains.

Wikipedia has struck a unique balance between open editing and the ability to rapidly detect and respond to vandalism. In 2005, Wikipedia gained a good deal of notoriety. A vandal posted false information that Kennedy journalist John Seigenthaler was involved in John F. and Robert Kennedys' assassinations. No one detected this error for several months. As a result of the Seigenthaler incident, Wikipedia improved their error detection, and later research demonstrated that erroneous information was corrected within hours of being entered.[121]

Wikipedia is often highlighted as a prime example of crowdsourcing success. Clay Shirky estimated that over 100 million hours of volunteer contributions have gone into Wikipedia,[122] the world's largest encyclopedia. During the course of its development, Wikipedia's user community has discussed and tested options for encouraging user contributions, while

---

[116] "Wikipedia.org Site Info," *Alexa*, n.d., http://www.alexa.com/siteinfo/wikipedia.org.

[117] Anselm Spoerri, "What Is Popular on Wikipedia and Why?," *First Monday* 12, no. 4 (April 2, 2007), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1765%E5%AF%86/1645.

[118] "Index of Geography Articles," *Wikipedia, the Free Encyclopedia*, November 17, 2011, http://en.wikipedia.org/wiki/Index_of_geography_articles.

[119] "List of Cities in Afghanistan," *Wikipedia, the Free Encyclopedia*, April 16, 2012, http://en.wikipedia.org/wiki/List_of_cities_in_Afghanistan.

[120] "Category:Lists of Mountains," *Wikipedia, the Free Encyclopedia*, February 10, 2012, http://en.wikipedia.org/wiki/Category:Lists_of_mountains.

[121] Brock Read, "Can Wikipedia Ever Make the Grade?," *Chronicle of Higher Education* 53, no. 10 (October 27, 2006): 27.

[122] Clay Shirky, *Cognitive Surplus: Creativity and Generosity in a Connected Age* (Penguin Press HC, The, 2010).

maintaining a reasonable level of quality and protecting against vandalism.

Figure 16. Wikipedia Infobox for Mount Everest[123]

Searching

Search tasks have long employed volunteers scouring the landscape to locate a particular individual, object, or feature of interest. When conducting searches in the distant past, individuals had to be physically present in the search area. More recent imagery-based searches had been limited in their effectiveness due to restricted access to imagery and image processing equipment and software.

This limitation changed dramatically with the widespread availability of imagery and web-based tools to make the imagery available to anyone with Internet access. Image-based searches could be crowdsourced by dividing imagery into small tiles and letting masses of volunteers scan the imagery for the items of interest. Thus, social media could be used to connect individuals looking for objects across wide areas.

While none of these techniques replace traditional search and rescue efforts, they offer new opportunities to bring additional resources to bear on the problem. Image-based search originated with the unsuccessful attempts to locate Microsoft executive Jim Gray,[124] who was lost sailing off the coast of California in 2007. Image-based searching was also

[123] "Mount Everest," *Wikipedia, the Free Encyclopedia*, May 15, 2012, http://en.wikipedia.org/wiki/Mount_Everest.

[124] Joseph M. Hellerstein and David L. Tennenhouse, "Searching for Jim Gray: a Technical Overview," *Commun. ACM* 54, no. 7 (July 2011): 77–87.

used in attempts to locate aviator Steve Fossett, who died in an airplane crash near Yosemite National Park in California in 2007.[125] Gray was never found and a hiker eventually found Fossett's body. Both the Gray and Fossett searches employed Amazon's Mechanical Turk,[126] a tool for crowdsourcing Human Intelligence Tasks (HITS). In each case, imagery covering a large area was tiled into individual images that multiple volunteers viewed and evaluated.

This section highlights two efforts, the National Geographic's Field Expedition: Mongolia - Valley of the Khans Project and the Defense Advanced Research Project Agency's Network challenge.

Field Expedition: Mongolia[127]

Field Expedition: Mongolia - Valley of the Khans Project is a National Geographic project focused on locating the tomb of Genghis Khan, as well as other cultural heritage sites in Mongolia. The project employed on-the-ground analysts who receive input from a number of high technology, non-invasive tools, including unmanned aerial vehicles (UAV), ground penetrating radar, and crowdsourced imagery analysis (Figure 17).

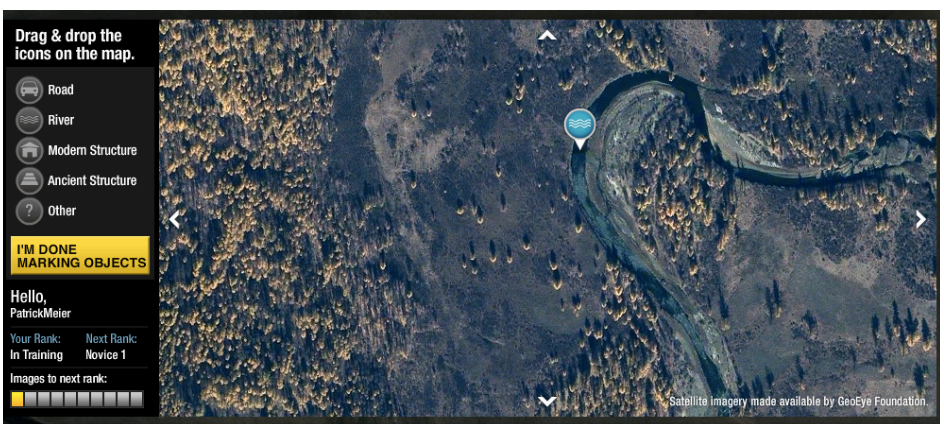


Figure 17. User interface for Field Expedition: Mongolia image analysis

Dr. Albert Yu-Min Lin, the project leader, noted that crowdsourcing was adopted for this task over automated image analysis, because it was not

---

[125] Kenneth Barbalace, "Internet Search for Steve Fossett Eight Weeks Later," *Yahoo! Groups*, October 31, 2007, http://mx.dir.groups.yahoo.com/group/Rescate/message/8050?var=1.

[126] "Amazon Mechanical Turk - Welcome," *Amazommechanical Turk*, n.d., https://www.mturk.com/mturk/welcome.

[127] "Field Expedition: Mongolia," *National Geographic*, n.d., http://exploration.nationalgeographic.com/.

possible to describe the exact nature and signatures of the features in advance. The goal was to have volunteers use a simple classification scheme to tag features in the images. It was the overall pattern, not any single observation that was important in focusing the attention of the scientists.

Also, as part of the process for this crowdsourcing search initiative, volunteers are informed about previous results for the same image, which is used as a catalyst to gain what Dr. Lin called "collective intelligence."[128] Over time he hopes to have everyone thinking and viewing features in a similar manner. The Mongolia project is one of the few examples of volunteer feedback found among the projects examined in this study.

The project was carried out using Tomnod crowdsourcing technology,[129] created to deal with the exponentially growing size and complexity of digital data sets. Tomnod focuses their efforts on data improvement, machine learning/automated computation, and crowd ranking technologies.

The information was evaluated and used to direct field observers in real-time. The results demonstrate the power of crowdsourced analysis. In the first two months, users classified 1.18 million features, while the effort resulted in the discovery of 55 previously unknown ancient burial sites. By tailoring the task to the user population, National Geographic was able to both obtain useful information and stimulate interest in their research.

DARPA Network Challenge[130]

In contrast to Field Expedition" Mongolia's imagery based search, the Defense Advanced Research Projects Agency's (DARPA) Network Challenge, also informally known as the Red Balloon Challenge, was a real-world search for the locations of ten weather balloons placed in public, but undisclosed locations around the United States (see Figure 18 and Figure 19).

---

[128] Ibid.

[129] "Tomnod: Crowdsource the World," *Tomnod: Crowdsource the World*, n.d., http://tomnod.com/.

[130] "DARPA Network Challenge," n.d., http://archive.darpa.mil/networkchallenge/.

Figure 18. Balloon locations during the DARPA competition[131]

Figure 19. Balloon #1 displayed in Union Square, San Francisco, CA[132]

This trial was a crowd wisdom challenge, where DARPA's real interest was to "explore the roles the Internet and social networking play in the timely communication, wide-area team building, and urgent mobilization required to solve broad-scope, time-critical problems."[133] About 4,300 teams participated in the challenge. The Massachusetts Institute of Technology (MIT) Red Balloon Challenge Team won the $40,000 prize. Their approach

... emphasized both speed (in terms of number of people recruited) and breadth (covering as much U.S. geography as possible). They set up a

[131] *Map*. PNG Image, 819 x 480 pixels, n.d. http://archive.darpa.mil/networkchallenge/BalloonMap.aspx

[132] *Balloon14*, JPEG Image, 2448 × 3264 pixels - Scaled (7%), n.d., http://archive.darpa.mil/networkchallenge/Photos/Balloon14.jpg.

[133] "DARPA Network Challenge | FAQ," n.d., http://archive.darpa.mil/networkchallenge/FAQ.aspx.

> platform for viral collaboration that used recursive incentives to align the public's interest with the goal of winning the challenge.[134]

Tracking

Tracking has its historical roots in the tracing of animal migration patterns. From first attempts to physically mark animals and record their sittings, to tagging and radio tracking, the goal has been to trace the movement of animals over time. This has been applied in other areas as well, such as law enforcement applications that track people and vehicles. With the advent of GPS technology, it is now possible to collect location and tracking information using low cost receivers found in special purpose devices or smartphones.

This section focuses on Waze, a site that combines information and uses tracks in multiple ways, from extracting speed information for traffic reporting, to identifying new roads and changes to existing roads, to sharing of favorite routes.

Waze[135]

Waze is a free social mobile network that provides users with traffic and road information in real-time. The commercial application is based on volunteered contributions, which are used to create traffic reports, update the underlying road database, and share favorite route information. Participation can be both passive and active.

According to the company, Waze works ...

> By simply driving with the app open on your phone, you passively contribute traffic and other road data that helps the Waze system to provide other Waze drivers with the most optimal route to their destination, including live traffic information. But you can also take a more active role by reporting on accidents, police traps, or any other hazards along the way, helping to give other users in the area a 'heads-up' about

[134] John C. Tang et al., "Reflecting on the DARPA Red Balloon Challenge," *Communications of the ACM* 54, no. 4 (April 1, 2011): 78–80.

[135] "Waze - Social Traffic & Navigation App," n.d., http://www.waze.com/.

> what's to come and contributing to the common good out there on the road.
>
> Some of the Waze community members with a passion for maps also take an even more active role by editing and updating the Waze map, itself. Most of the editing work is done on the Waze website, but some parts, such as the naming of streets, can be done through the application directly.[136]

Like StreetBump,[137] Waze incorporates the passive collection of content, where users simply turn on the application and data is automatically sent to the Waze servers. A feature of interest from the crowdsourcing perspective is the synthesis of user inputs for real time traffic updates and addition of roads. Road traffic updates are triggered by contributions from two or more users. For road network updates, Waze notes that "between 20 and 100 trips accurately recorded seem to be enough to trigger Waze to make an automatic update to the roads."[138]

The data utilized by Waze is based on publicly available TIGER[139] data from the US Census Bureau and is updated by Waze users, either automatically through driving or via a map editor. User requests for incorporating OSM[140] data have been rejected due to OSM licensing models, which may inhibit future business use of the data within Wave.[141]

Participation of a large number of volunteers is essential to the success of the Waze, which relies on continuous reporting over large areas to provide timely and accurate travel information.

Validating

Once geospatial data is collected, it requires validation to ensure the overall quality of the content. Validation usually involves the assessment of

136 "Waze: Way to Go," n.d., http://www.waze.com/faq/#6.

137 "Street Bump."

138 "Timeline of Updating Process," *Waze*, February 12, 2012, http://www.waze.com/wiki/index.php/Timeline_of_updating_process.

139 US Census Bureau Geography Division, "US Census Bureau TIGER/Line," n.d., http://www.census.gov/geo/www/tiger/shp.html.

140 "OpenStreetMap."

141 "Waze: Way to Go."

positional, attribute, and topological accuracy. While quality assurance is best performed at the time of data entry, post-production validation remains an essential task.

Validation lends itself to crowdsourcing applications, where multiple looks at data can identify anomalies. This is the foundation of the Linus's Law, attributed to Eric Raymond, author of 'The Cathedral and the Bazaar,' which in referring to software development states that "given enough eyeballs, all bugs are shallow" or more formally "Given a large enough beta-tester and co-developer base, almost every problem will be characterized quickly and the fix will be obvious to someone."[142]

One well-known, non-geospatial application in this area is the Australian Newspapers Digitisation Program, sponsored by the Australian National Library.[143] In this initiative, volunteers review and correct Optical Character Recognition (OCR) text extracted from Australian newspapers. This work has resulted in corrections to 12.5 million lines of text by over 9,000 volunteers.

Four geospatial validation applications are highlighted in this section: NAVTEQ Map Reporter, the GeoWiki Project, Old Weather, and OSM. Each application represents a different approach to validation.

NAVTEQ Map Reporter[144]

NAVTEQ is the leading provider of navigation data, including maps, traffic, and location data.[145] Their data is used in the automotive industry, for fleet and logistics, the Nokio Maps Internet map service, and by Government agencies. NAVTEQ accomplishes this task by "tapping over 80,000 sources, but going beyond that when necessary to the quality of the experience and putting approximately 1,100 geographic analysts around the world in the field to collect the right data."[146]

---

[142] Eric S. Raymond, "Release Early, Release Often," *The Cathedral and the Bazaar*, August 2, 202AD, http://www.catb.org/~esr/writings/homesteading/cathedral-bazaar/ar01s04.html.

[143] "Australian Newspapers Digitisation Program," *National Library of Australia*, February 17, 2012, http://www.nla.gov.au/ndp/.

[144] "NAVTEQ Map Reporter," n.d., http://mapreporter.navteq.com/.

[145] "NAVTEQ Corporate - About Us," n.d., http://corporate.navteq.com/.

[146] Ibid.

Despite their massive data collection efforts, NAVTEQ recognizes that imperfections exist within their data, and utilizes the NAVTEQ Map Reporter application to collect corrections and updates from product users (See Figure 20).

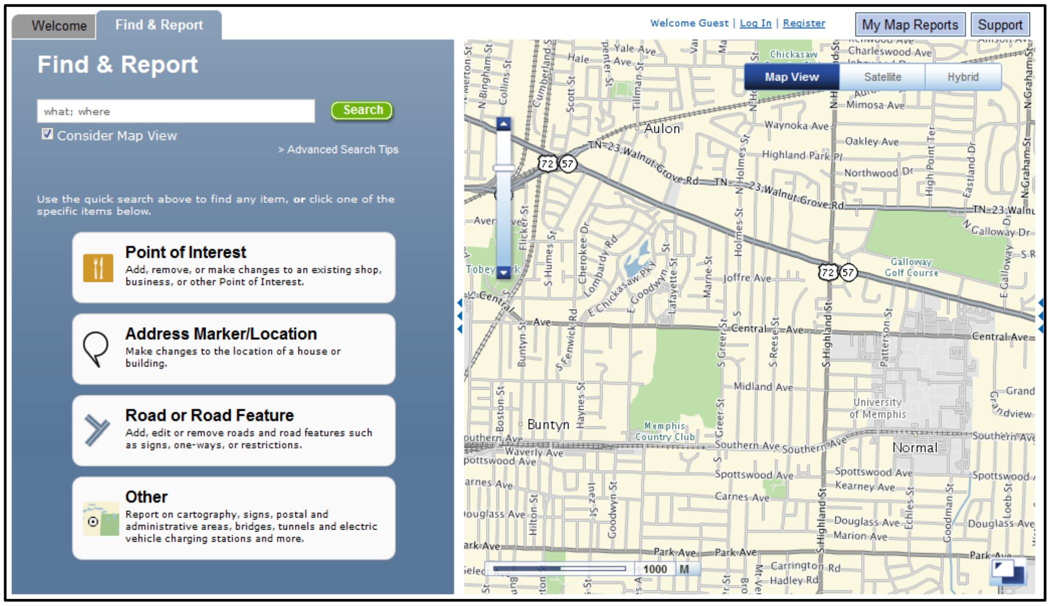


Figure 20. NAVTEQ map reporter application[147]

Contributor-proposed edits are evaluated before making changes to the database. Using a rules-based system, some edits are automatically accepted, while others are sent to field teams for verification. NAVTEQ Map Reporter is a good example of a hybrid crowdsourcing model, where the crowd contributes changes that are vetted and approved by an authoritative mapping organization for incorporation in their proprietary product.

The Geo-Wiki Project[148]

The Geo-Wiki Project taps into an international network of volunteers to address quality issues in global land cover mapping by calculating the differences among three major global land cover products, GLC-2000,[149] MODIS,[150] and GlobCover.[151] Contributors validate the quality of classifi-

---

147 "NAVTEQ Map Reporter."

148 "The Geo-Wiki Project," *Geo-Wiki*, n.d., http://geo-wiki.org/login.php?ReturnUrl=/index.php.

149 "Global Land Cover 2000 | GEM - Global Environment Monitoring," n.d., http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php.

150 "MODIS," n.d., http://modis.gsfc.nasa.gov/.

151 "GlobCover | ESA," n.d., http://due.esrin.esa.int/prjs/prjs68.php.

cations in hotspot areas using high-resolution imagery, ground photography, and their local knowledge. Their input is collected in a database that will contribute to future land cover mapping (Figure 21).

Geo-Wiki is an example of a project that requires subject matter expertise, rather than tapping common knowledge. In order to evaluate global land cover, users must be familiar with the three global land cover products terminology, as well as a basic understanding of land cover. This limits the pool of individuals who can successfully contribute to the project, but is typical of more specialized crowdsourcing applications.

Although guests can access the system, volunteers must be registered to enter data. When providing responses, users can view the footprints and selected class of the different global land cover maps against a background Google Maps imagery. They register an opinion about the quality of the classification (Good, Not Sure, Bad) and are given an opportunity to select a more appropriate classification if they feel the data is misclassified, as well as provide confidence in their estimates. To assist in the classification estimate, volunteers may access ground photography from the Confluence Map Project[152] or Panoramio.[153]

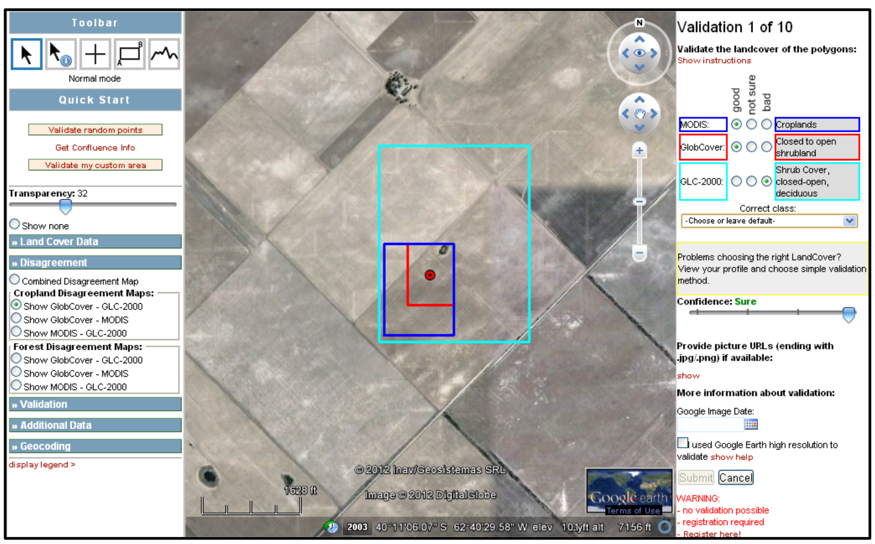


Figure 21. Image of Geo-wiki project Web page[154]

[152] "The Degree Confluence Project," n.d., http://confluence.org/.

[153] "Panoramio - Photos of the World," n.d., http://www.panoramio.com/.

[154] "The Geo-Wiki Project."

OSM Inspector[155]

OSM Inspector, one of a number of OSM validation tools, displays potential errors relating to geometry, tagging, and the route network. Figure 22 displays a view of geometry issues over the Mid-Atlantic region. To ensure confidence in the quality of the product, errors are explicitly identified to anyone viewing the map, and can be corrected by the multitudes of OSM contributors. This open format grants collaborative quality-control abilities not featured in authoritative data products.

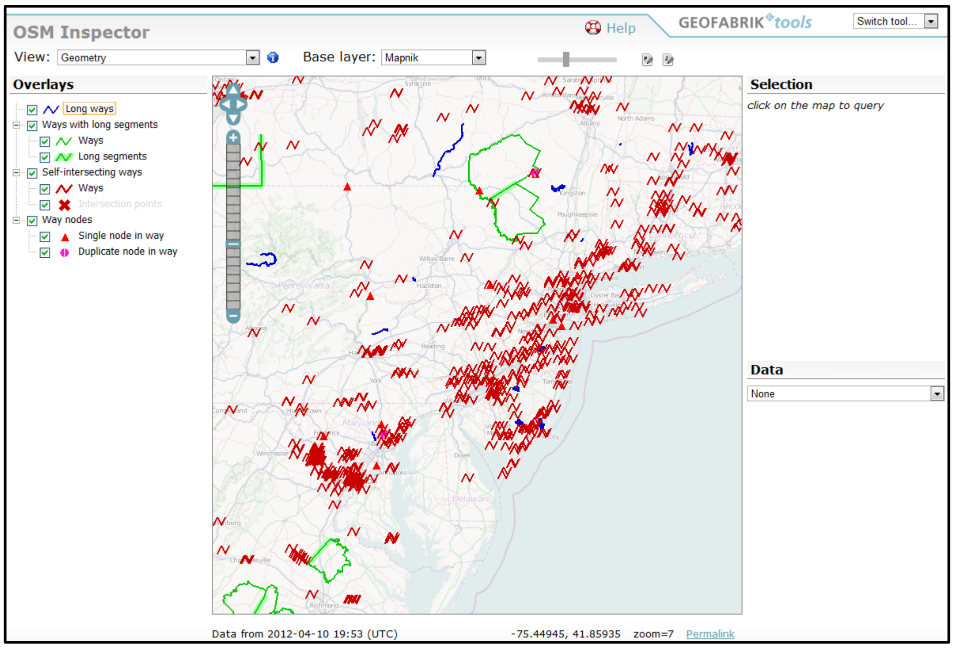


Figure 22. OSM Inspector view of geometry errors in the Mid-Atlantic region

OSM Inspector offers an overview of errors, while allowing users to zoom in, inspect, and correct specific issues. In Figure 23, a self-intersecting parking lot outline is shown in the center. By directing contributors to known problems and potential problems in the data, OSM Inspector taps the power of its large user base to improve the overall quality of the data.

155 "OSM Inspector," *Geofabrik Tools*, 2011, http://tools.geofabrik.de/osmi/.

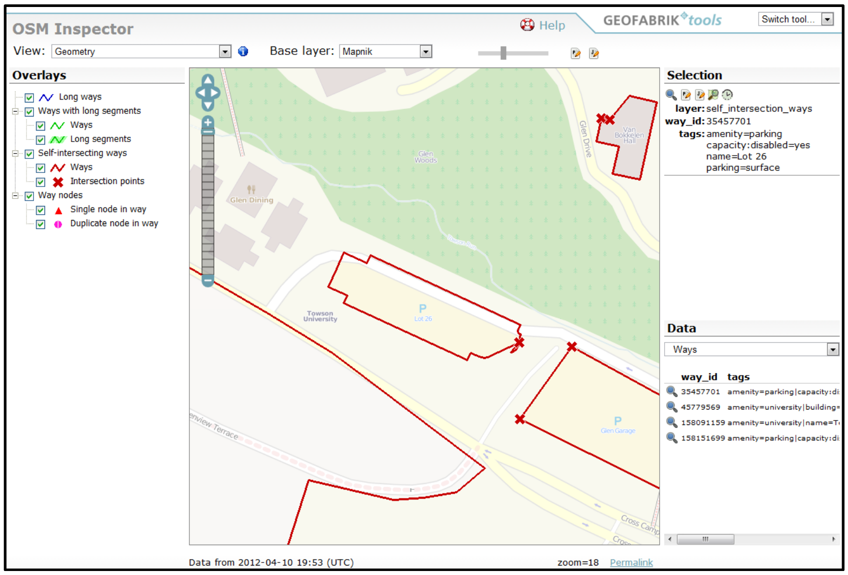


Figure 23. OSM Inspector view of individual error

Polling/Surveying

James Surowiecki's book, "The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations," begins with a story about crowd polling at a country fair in the late 1800's.[156] Francis Galton was looking to show the value of expertise over common knowledge and was greatly surprised to find that an averaged crowd estimate was closer to the truth than single estimates from experts.[157] While polls are not applicable in all situations, they are effective when diverse collections of individuals weigh in on problems not requiring specific technical expertise.

Because Web 2.0 technology allows organizations to quickly create, distribute, and analyze responses in real time, polling and surveying technologies are more widely available. Leading competitors in this market include SurveyMonkey,[158] Google Docs Forms,[159] SurveyGizmo,[160] and Zoomerang.[161] All of the above products are sophisticated text-based solu-

---

[156] James Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations* (Doubleday, 2004).

[157] Ibid.

[158] "SurveyMonkey: Free Online Survey Software & Questionnaire Tool," *SurveyMonkey*, n.d., http://www.surveymonkey.com/.

[159] "Google Docs," *Google Docs*, n.d., http://www.google.com/google-d-s/forms/.

[160] "Online Survey Software | SurveyGizmo - Advanced Survey Software," *Surveygizmo*, n.d., http://www.surveygizmo.com/.

[161] "Online Survey Software - Create Online Surveys," *Zoomerang*, n.d., http://www.zoomerang.com/.

tions. SurveyMapper, however, has a geospatial component that the other applications are missing.

SurveyMapper[162]

SurveyMapper is a free real-time geographic survey and polling tool. Unlike text-based survey tools, SurveyMapper integrates place-based information in surveys and displays survey results on a map.

SurveyMapper supports a range of place names, including the following: country names, United States state names, United States ZIP codes, United Kingdom counties, United Kingdom postcodes, London boroughs, and London wards. In addition, worldwide markers are also used. Creating a survey is a process of providing basic information about the survey, including a locational reference (see Figure 24).

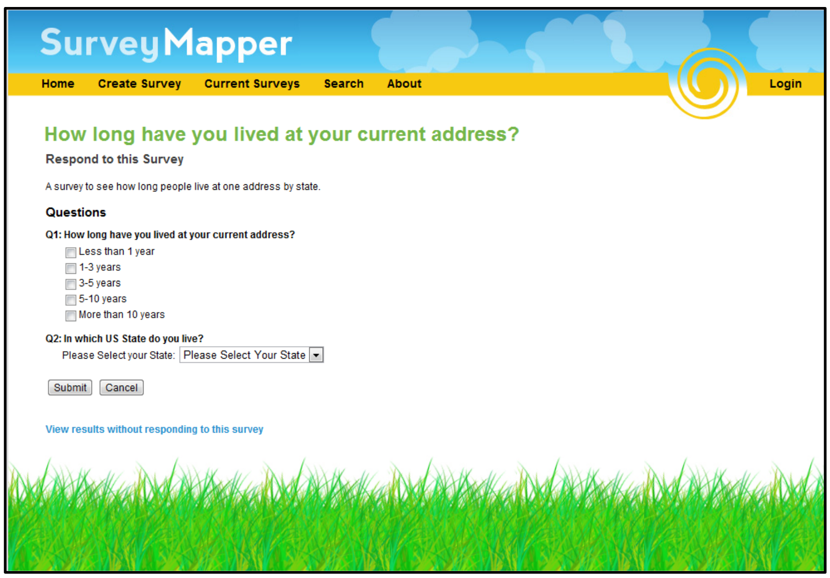


Figure 24. SurveyMapper response form. The location component of the questionnaire on the topic of 'How long have you lived at your current address?' is Question 2, where the respondent identifies their state of residency

Place-based entries are automatically linked to their related features on maps, and the results are displayed graphically (see Figure 25). The map database is updated as soon as a response is received and displays the results in real-time. Each survey has an associated analytics page that dis-

162 "Welcome to SurveyMapper," *SurveyMapper*, n.d., http://www.surveymapper.com/.

plays responses over time, charts for each question, and a list of the top places.

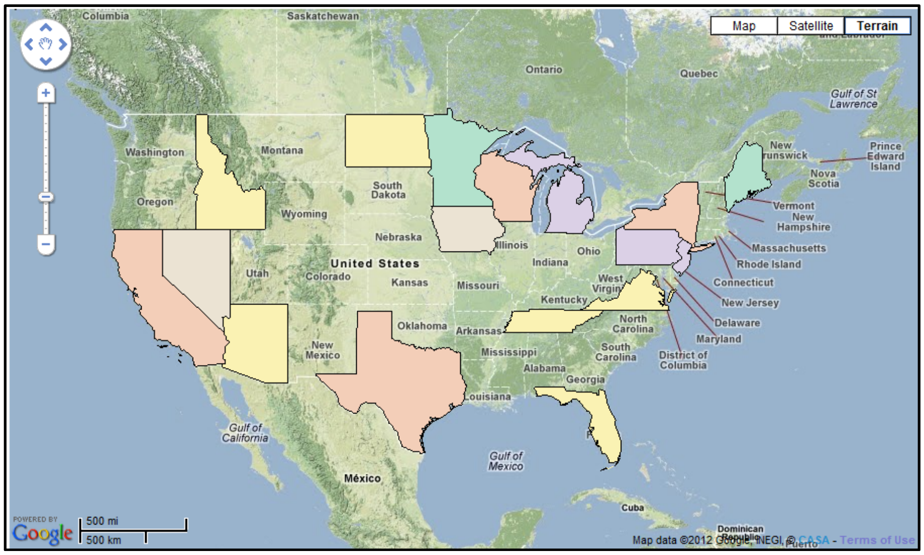


Figure 25. Map of results to survey 'How long have you lived at your current address?'

Socializing

The development of Web 2.0 technologies has resulted in a wide range of tools to facilitate online interaction that supports the growth of online communities. Social media tools were introduced in the 1990's and expanded greatly after 2000. Services such as blogs, wikis, social bookmarking, social networks, and media sharing services emerged; engaging large segments of the population.

While social media sites collect georeferenced and time-stamped information, their main goal is not to create geospatial databases. They do, however, store a tremendous amount of information that, when properly organized and analyzed, provides valuable geospatial data.

Currently, users access social media websites through smartphones and the Web. This convenient accessibility allows users to provide up-to-date, geospatial information anytime and anywhere. Three popular social sites are reviewed: Twitter, Flickr, and Foursquare.

Twitter[163]

Twitter is a free open source social network site that provides users with real-time information[164] through short, 140 character messages that answer the question 'What's happening?'

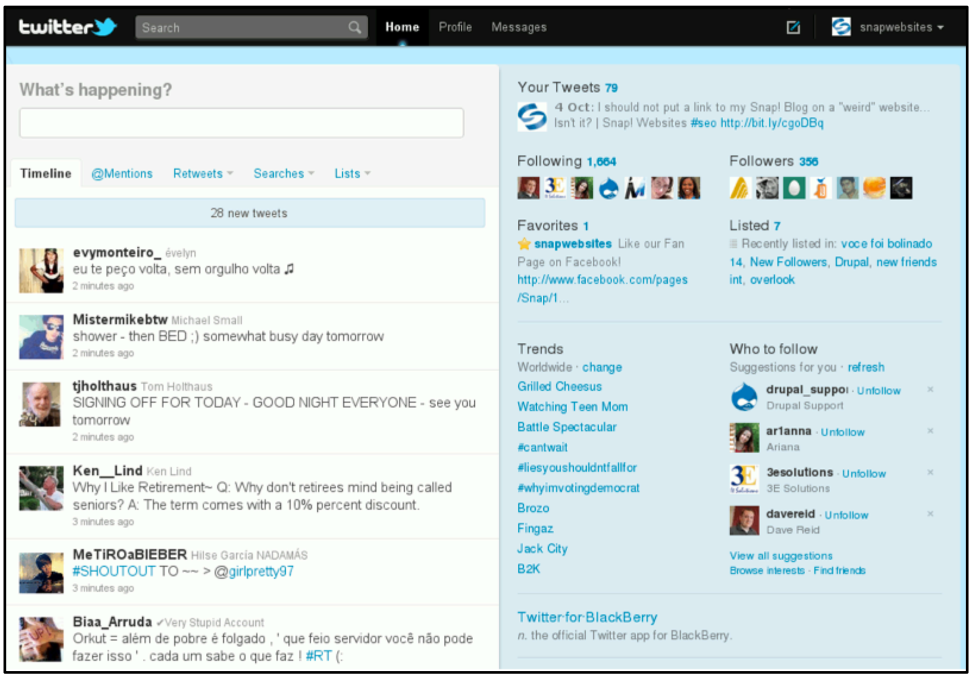


Figure 26. Screenshot of Twitter homepage of a registered user[165]

There are three types of geospatial location information that may be associated with a Twitter message or tweet: the place information provided in the user profile, the location from where the message was tweeted, and the places mentioned in the tweet. Previous research has shown that roughly 66% of the user profiles have valid locations entered, while less than 12% of tweets record the location from where they were tweeted. [166]

While each message may contain only a small amount of information, the large volume of traffic can be analyzed for trends and mapped. Twitter has

---

[163] "Welcome to Twitter," *Twitter*, n.d., https://twitter.com/.

[164] "Twitter: About," *Twitter*, April 7, 2012, https://twitter.com/about.

[165] "New Twitter - A New Look and Functionality," *Snap! Websites Journal*, October 2010, http://snapwebsites.info/journal/2010/10/new-twitter-new-look-and-functionality.

[166] B. Hecht et al., "Tweets from Justin Bieber's Heart: The Dynamics of the Location Field in User Profiles," in *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 2011, 237–246.

been used to map everything from the weather,[167] to the mood of a nation;[168] and has demonstrated the ability to provide valuable, time-critical, geospatial user information. Twitter messages have been used to disseminate information about the Santa Barbara's wildfires,[169] the Red River floods and Oklahoma grass fires,[170] as well as the terrorist attacks on Mumbai.[171]

Flickr[172]

Flickr, from Yahoo! Inc., is a media sharing site that allows users to manage and share photos and videos online[173] (Figure 27). As of September 2010, the site had over 5 billion photos.

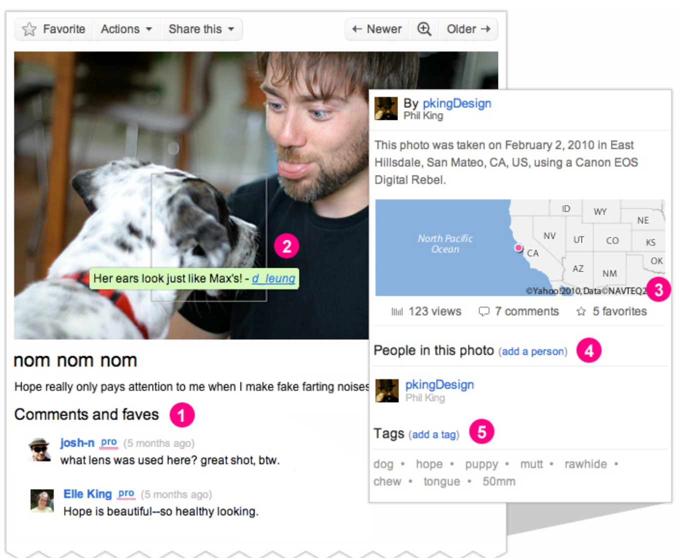


Figure 27. Description of features: 1) Comments made by other users, indicating profile name and date of posting, 2) Notes or comments on the picture, 3) Mark as favorite for easy access later, 4) Tag people in pictures, and 5) Categorize pictures[174]

---

167 "Social Weather Mapping with Twitter," *Information Aesthetics*, March 19, 2009, http://infosthetics.com/archives/2009/03/social_weather_mapping.html.

168 Celeste Biever, "Twitter Mood Maps Reveal Emotional States of America," *NewScientist*, July 21, 2012, http://www.newscientist.com/article/mg20727714.200-twitter-mood-maps-reveal-emotional-states-of-america.html

169 Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

170 S. Vieweg et al., "Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness," in *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, 2010, 1079–1088.

171 Onook Oh, Manish Agrawal, and H. Raghav Rao, "Information Control and Terrorism: Tracking the Mumbai Terrorist Attack Through Twitter," *Information Systems Frontiers* 13, no. 1 (September 25, 2010): 33–43.

172 "Welcome to Flickr!," *Flickr*, n.d., http://www.flickr.com/.

173 "About Flickr," *Flickr*, n.d., http://www.flickr.com/about/.

174 "Tell a Story with Your Photos," *Flickr*, n.d., http://www.flickr.com/tour/#section=tell-a-story.

Photos may be geotagged in a number of different ways (see Figure 28). Many cameras are GPS-enabled, enabling them to record the location of a photograph directly in the Exchangeable Image File Format (EXIF) file associated with the image. The EXIF file can also be updated with GPS co-ordinates from a separate device or by associating the photograph with a map location. Place information can also be added through photo tags. Mapping applications typically access the coordinate or tag information and collections of photos can be used to identify geographic features.[175]

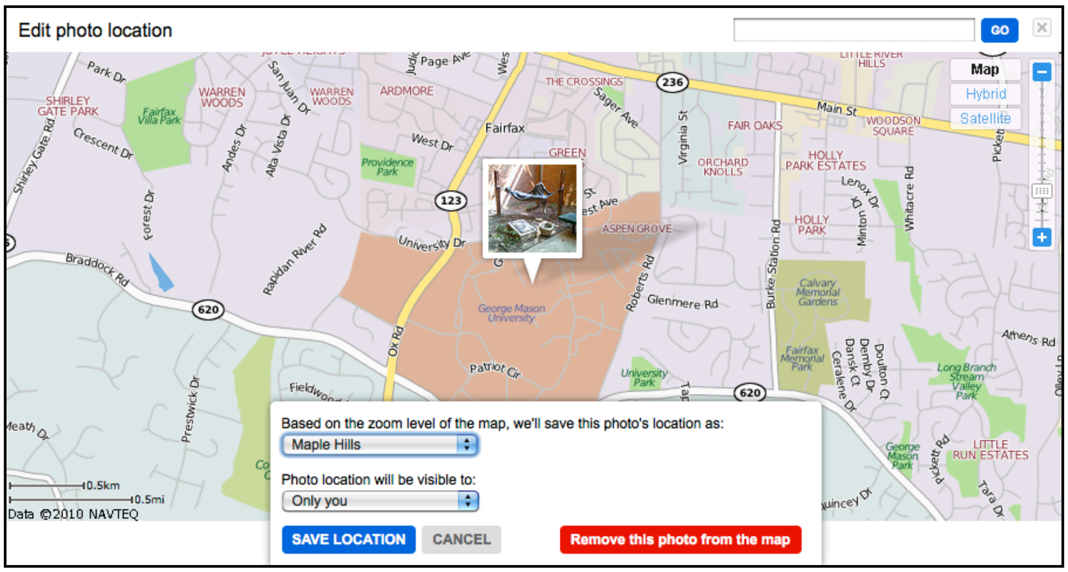


Figure 28. User interface for adding the location of an uploaded picture

Foursquare[176]

Foursquare is an open-source web and mobile application that enables interaction between users interested in finding local hotspots by allowing them to share comments about places and providing their current location. The site currently has over 20 million members and over two billion check-ins.

[175] Aaron, "The Shape of Alpha," *Code: Flickr Developer Blog*, October 30, 2008, http://code.flickr.com/blog/2008/10/30/the-shape-of-alpha/.

[176] "Foursquare," *Foursquare*, n.d., https://foursquare.com/.

Users are able to leave comments about locations, events and promotions, and then share them with friends through the Foursquare application on a mobile device or via a web site. Figure 29 provides a screenshot of the "check in" interface of the Foursquare's mobile application.

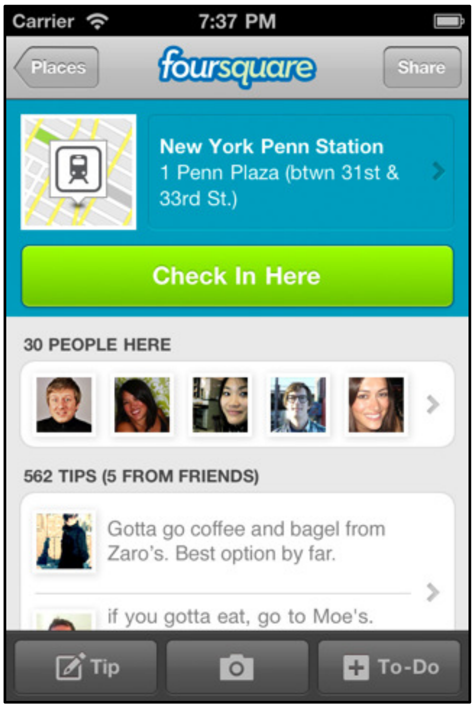


Figure 29. Foursquare location "check in" screen[177]

A few weeks after launching the application, Foursquare received reports of fraudulent location submissions. Some dishonest users found a way to take advantage of the Foursquare's system to gain more rewards, such as discounts, and free products, or just have fun.[178] Mai Ren conducted an investigation of the methods and causes of fraud on Foursquare.[179] Ren describes four ways of commiting fraud in Foursquare: 1) manipulate mobile devices to provide fake GPS information, 2) crawling data from Foursquare's website, 3) automated cheating, and 4) cheating with venue profile analysis.[180]

Foursquare has developed "Cheater code" to address the problem of cheating, in "an attempt to catch some of the folks that are checking in from their couches to steal mayorships."[181] Figure 30 shows an example of the message that users get when the check-in venue contradicts the location provided by the GPS located in the users' mobile devices.

[177] "Foursquare," *iTunes Store*, March 27, 2012, http://itunes.apple.com/us/app/foursquare/id306934924?mt=8.

[178] Jessica Guynn, "Confessions of a Foursquare Cheater," *LATimes.com (blog)*, February 16, 2010, http://latimesblogs.latimes.com/technology/2010/02/confessions-of-a-foursquare-cheater.html.

[179] Mai Ren, "Location Cheating: A Security Challenge to Location-based Social Network Services," *Computer Science and Engineering: Theses, Dissertations, and Student Research* (December 1, 2011), http://digitalcommons.unl.edu/computerscidiss/31.

[180] Ibid.

[181] "On Foursquare, Cheating, and Claiming Mayorships from Your Couch...," *Foursquare (blog)*, April 7, 2010, http://blog.foursquare.com/2010/04/07/503822143/.

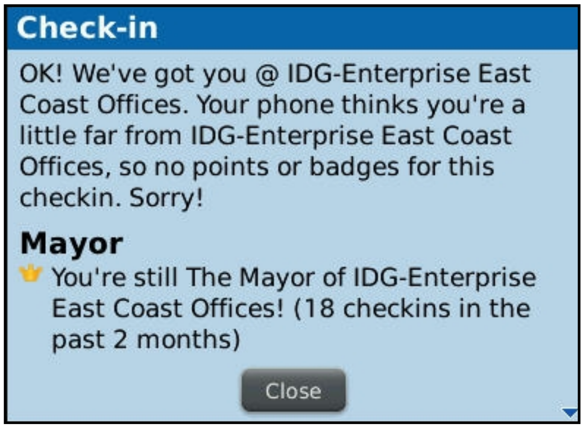


Figure 30. Foursquare "Cheater Code" Check-In Error Message[182]

While cheating on Foursquare is not widespread, it appears to be pervasive; brought about by the game-like nature of the application that causes participants to break the rules to gain an advantage. The lessons learned from Foursquare's experience are a cautionary note to other crowdsourced geospatial applications, particularly related to the integrity of locational information.

Sharing

Sharing sites host geospatial content placed by crowd members, including data, applications, or finished cartographic products. While socializing sites focus on non-geospatial content but have a geospatial component, this section focuses on sharing geospatial products and applications. The content, which often is stored in the cloud, is available to other users, who can access, repurpose, and visualize it.

Sharing sites typically offer data sharing capabilities well beyond the simple upload and download of data. They also include: tools to view the data, the ability to mash-up data with other content, customization of the user experience, and interaction with developers and other users through social media. Their goal is to make the sharing and visualization of data as simple as possible, while promoting a community of sharing.

ArcGIS Online and GeoCommons are two geospatial sharing sites focusing on sharing geospatial content. These sites not only offer upload and download of data along with access to web services, but they also incorporate analytical tools, advanced cartographic design tools, as well as tools to fuse and mash-up data with other content.

[182] "Foursquare Addresses Cheating Issue, Frustrates Legit Users," *CIO Blogs*, April 8, 2012, http://advice.cio.com/al_sacco/10000/foursquare_addresses_cheating_issue_frustrates_legit_users

ArcGIS Online[183]

ArcGIS Online, from Esri, allows users to "create, store, and manage maps, apps, and data, and share them with others. You also get access to content shared by Esri and GIS users around the world"[184] (see Figure 31). This cloud-based system is free to users, but focuses primarily on Esri-generated content. It is well positioned as a content management system.

ArcGIS Online content, including data, geoprocessing workflows, and maps can be uploaded to the site, where other users can discover it. ArcGIS features some online analytical capabilities, such as geocoding, and hosts web-based analysis services as well.

Data documentation does not rely on traditional metadata, but incorporates an abbreviated description of the information along with tags. This is a folksonomic approach similar to Flickr tags, which tends to be richer and more flexible than traditional classification systems. However, it also introduces noise as tags may not be standardized.

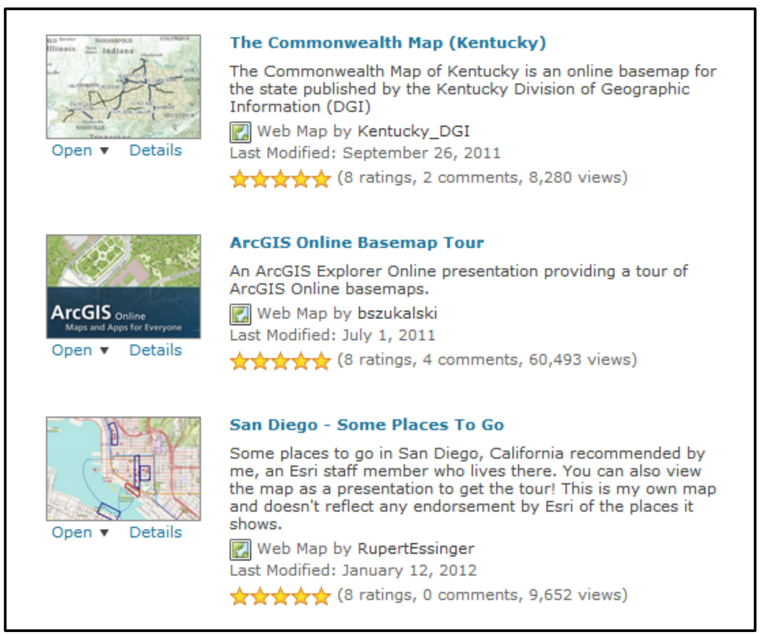


Figure 31. ArcGIS Online search results, showing thumbnail images, descriptions, ratings, comments, views, and capability to open map in multiple viewers

[183] "ArcGIS Online," *ArcGIS Online*, n.d., http://www.arcgis.com/home/.

[184] "Free Personal Account," *ArcGIS Online*, n.d., http://www.esri.com/software/arcgis/arcgisonline/features/free-personal-account.html.

GeoCommons[185]

GeoCommons is a “public community of GeoIQ users who are building an open repository of data and maps for the world.”[186] The main web site shares a similar look and feel with ArcGIS Online, but leverages the GeoIQ platform, rather than the Esri suite of software.

GeoCommons differs from other sharing sites with its inclusion of a broad range of analytical capabilities, such as merging, aggregating, buffering, filtering, clipping, intersecting, and performing various calculations. Users can perform geospatial analysis operations on their data before visualizing it.

A suite of visualization tools support manipulation of the visual variables in cartographic design, with the ability to change the size, shape, and color of symbols.

185 “GeoCommons.”

186 “A Tour Through GeoCommons,” *Geocommons*, n.d., http://geocommons.com/tour.

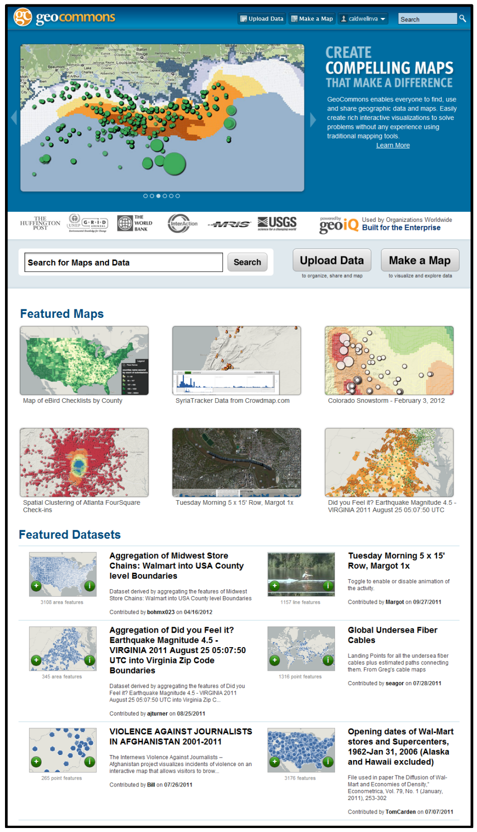


Figure 32. GeoCommons home page[187]

Summary

The examples in this chapter surveyed a wide range of approaches for CGD. The variations with which these applications exist exemplify their utility and inherent flexibility that allows them to fill existing data gaps and solve many current and future problems.

Table 4 through Table 8 (Appendix 2), summarize the survey, highlighting the key aspects of each project, including tasks addressed, geospatial data entry options, geospatial data geometries collected, and other relevant information. These tables may be particularly useful in comparing applications and activities associated with them. The tables are not intended to be a comprehensive survey of CGD applications or a comprehensive description of each application, but is included to facilitate comparisons and to present this chapter's information in a more condensed format.

187 "GeoCommons."

The following chapter deals with an important issue that has been widely discussed and considered by academics and geospatial practitioners: quality issues associated with CGD. This topic has a long history within the geospatial community and has been the subject of many outstanding research papers and practices, some of which we will present in summary form. In Chapter 4 we will review the main issues associated with quality and discuss CGD quality from the perspective of several well-known approaches in published literature.

4 Quality and Crowdsourced Geospatial Data

Introduction

In the previous three chapters, CGD concepts, production methods, and applications have been presented. This chapter addresses the important topic of quality in CGD.

Quality is undoubtedly one of the most important considerations in determining whether CGD and CGD-based production methods and tools should be adopted in a geospatial project or enterprise. Many examples of CGD come from non-profit groups, businesses, and the social media world, where considerations of quality are important, but perhaps less so than for government agencies with specific mandates and accountability for accuracy, quality, and reliability. Nevertheless, CDG quality should be an important consideration for all communities, regardless of the mandate they operate under.

Quality encompasses a number of topics and related terms, including: accuracy, lineage, completeness, consistency, temporality, reliability, robustness, truthfulness, and credibility. In order to provide an informative and manageable discussion about this very broad topic, we present the most relevant aspects of quality from a traditional point of view, and from a CGD point of view, and offer a number of references that the reader can consult for more information.

Traditional Data Quality Concepts

Traditional geospatial data quality concepts have been in development for a very long time, arguably since the Age of Discovery, when maps of new lands gained tremendous value to the nations of Western Europe. Accurate maps of the New World were elevated to the level of strategic state secret, and were considered to be among the most valuable resources.[188] As map production and distribution programs continued in the United States

[188] For an outstanding book on the history of cartography and map publishing see Mary Sponberg Pedley, *The commerce of cartography: making and marketing maps in eighteenth-century France and England* (Chicago, Illinois: University of Chicago Press, 2005).

during the late 19th and 20 centuries, notions of quality and accuracy began to formalize.

Many of the traditional concepts were initially based on the idea of a printed, paper map, but have subsequently been generalized and adapted by authors to electronic maps and GIS databases. These more recent quality concepts can be extended to CGD. The basic concepts we will discuss, as they relate to CGD, are lineage, positional accuracy, attribute accuracy, temporal quality, logical consistency, and completeness.

Lineage

In the Federal Geographic Data Committee's (FGDC) "Content Standard for Digital Geospatial Metadata, lineage is defined as "information about events, parameters, and source data which constructed the data set, and information about the responsible parties."[189] Ideally, lineage is stored as a component for each feature in the database, and presents a complete history of the source data, including data collection events, processing, provenance, and custodianship.

Lineage is especially important for CGD applications, where contributions may come from a number of different sources of varying quality. OSM allows users to import existing geospatial data, enter GPS coordinates, or digitize from maps and imagery. With Twitter, users can enter locations using place names or have their location automatically calculated based on Internet Provider geolocation or GPS. CGD often differs from authoritative mapping production, which is typically constrained to a limited number of approved data sources.

Heinis and Alonso (2008) assert that "not knowing the exact provenance and processing pipeline used to produce a derived data set often renders the data set useless from a scientific point of view."[190] They note that modern workflow tools are better at capturing and preserving lineage than earlier tools, yet more efficient methods are needed to use the preserved lineage information.

---

[189] "Geospatial Metadata Standards — FGDC Endorsed ISO Metadata Standards," *Federal Geographic Data Committee*, April 25, 2012, http://www.fgdc.gov/metadata/geospatial-metadata-standards#fgdcendorsedisostandards.

[190] Thomas Heinis and Gustavo Alonso, "Efficient Lineage Tracking for Scientific Workflows," in *Proceedings of the 2008 ACM SIGMOD International* (presented at the Conference on Management of Data, ACM Press, 2008), 1007–1018, http://dl.acm.org/citation.cfm?id=1376716.

Although lineage is a critical metadata element, and a required part of the FGDC metadata standard, lineage is often not present in CGD. Girres and Touya's 2010 study of OSM data in France show that only 27.8% of the objects sampled in their study contained information about data source, and only 6.0% contained information about software. They suggest that lineage information would be a useful way to mediate contributions from non-authoritative sources and improve the quality of CGD.[191]

Positional Accuracy

As perhaps the best-known quality issue, positional accuracy has been explored by a number of mapping organizations and researchers for decades. Accuracy was the very first topic for intensive research during the 20-year research program of the National Center for Geographic Information and Analysis.[192] This research initiative resulted in a number of papers and books on analyzing, improving, and dealing with error in GIS, notably Goodchild and Gopal's 1989 work, *Accuracy of Spatial Databases*.[193]

In its most basic form, positional accuracy refers to the deviation of mapped feature positions from their actual positions in the horizontal and vertical domains. For printed and fixed-scale maps, a positional accuracy standard was developed in the early 1940s and published in 1947 as the US National Map Accuracy Standards (NMAS).[194] This standard, while useful for printed maps and as a basis for simple heuristics to estimate positional accuracy at a given scale, is not used in modern geospatial applications where printed maps are uncommon and scales can change. For applications involving geospatial data at multiple scales, the National Map Accuracy Standards have been replaced with the National Standard for Spatial Data Accuracy (NSSDA), which uses a statistical methodology for estimating the positional accuracy of maps and geospatial data.[195] Both standards will be briefly presented, followed by a general discussion of positional accuracy and CGD.

---

[191] Girres and Touya, "Quality Assessment of the French OpenStreetMap Dataset."

[192] "NCGIA Research Initiatives," *NCGIA*, n.d., http://www.ncgia.ucsb.edu/research/initiatives.html.

[193] Michael F. Goodchild and Sucharita Gopal, *The Accuracy of spatial databases* (London; New York: Taylor & Francis, 1989).

[194] "National Geospatial Data Standards - United States National Map Accuracy Standards," *USGS*, October 28, 2011, http://nationalmap.gov/standards/nmas.html.

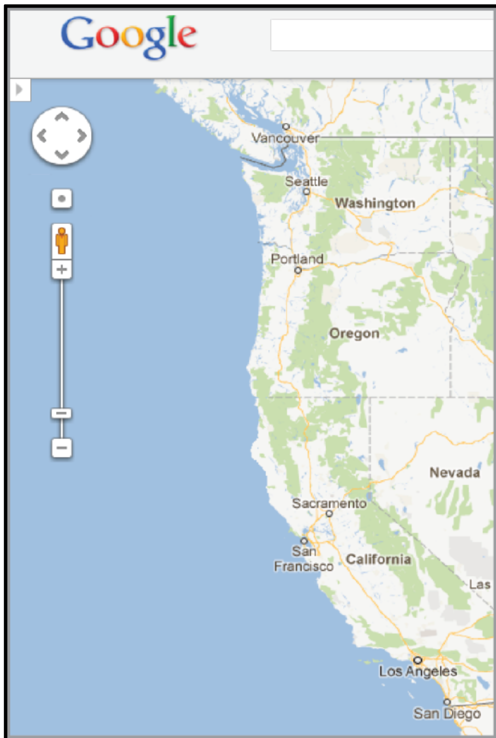
[195] T. V. Authority, "Geospatial Positioning Accuracy Standards Part 1: Reporting Methodology" (National Aeronautics and Space Administration, 1998), http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/chapter3.

The NMAS state that for published maps, 90% of well-defined features sampled should have horizontal positional errors of less than 1/30th of an inch at publication scale (for maps published at 1:20,000 or larger, the figure is reduced to 1/50th of an inch).[196] This suggests that for the standard US Geological Survey 7.5' topographic quadrangle map published at 1:24,000, 90% of the well-defined features sampled for positional accuracy should be within 40 feet, or approximately 12 meters. Vertical errors, according to the NMAS, are to not to exceed ½ the elevation contour interval for the same 90% sample.

The NSSDA has a different conceptual basis than the NMAS. It was designed by the FGDC to reflect the presence of geospatial databases and interactive mapping applications that are not constrained to a single, fixed scale, as is the case with a printed map and the NMAS. For devices such as GPS and smartphones (which are commonly used in CGD collection), the NSSDA is more appropriate.

To illustrate the wide ranging scales (and spatial resolutions) associated with CGD, Figure 33 is provided as a reference. Two of the most common current web mapping applications (Google Maps and Microsoft Bing Maps), allow users to zoom to 25 different scales, with associated pixel resolutions ranging from 157 km (for a base map of the entire earth displayed in a computer window) down to 9.3 mm (for a similar map showing a very small section of a small neighborhood parcel) at the equator. In addition, resolution also varies as a function of latitude, adding to the complexity of determining the pixel resolution. The National Map Accuracy Standard from the 1940s was simply not designed to measure and characterize error in applications that change scales this significantly.

[196] "National Geospatial Data Standards - United States National Map Accuracy Standards."



Zoom Level	Pixel Size at the Equator
0	157 km
1	78 km
2	39 km
3	19.6 km
4	9.8 km
5	4.9 km
6	2.4 km
7	1.2 km
8	611 m
9	306 m
10	153 m
11	76 m
12	39 m
13	19 m
14	10 m
15	5 m
16	2.4 m
17	1.2 m
18	60 cm
19	30 cm
20	15 cm
21	7.5 cm
22	3.7 cm
23	1.9 cm
24	9.3 mm

Figure 33. Zoom levels and pixel size for Google Maps and Microsoft Bing Maps

For NSSDA, there is no positional accuracy threshold or scale-based criteria for conformance with the standard, as with the NMAS. Federal agencies that collect or produce geospatial data are encouraged to set their own criteria for acceptable accuracies, and report their accuracies according to the methodology outlined in NSSDA.

The NSSDA uses a common statistical error measure called root-mean-square error (RMSE), which is the square root of the average squared deviations of sampled points from a source of ground truth. The results of the NSSDA-based positional accuracy assessment are reported using a 95% confidence interval, which implies that less than 5% of observations will have a positional error greater than the reported error confidence limits.

Importantly, the NSSDA standard suggests that geospatial datasets may contain themes with different accuracies, and geographic areas with different accuracies. For CGD, and more particularly for hybrid projects where authoritative and asserted geospatial data are combined, this is very

likely, or even certainly to be the case. For these cases, the NSSDA suggests:

- If data of varying accuracies can be identified separately in a dataset, compute and report separate accuracy values.
- If data of varying accuracies are composited and cannot be separately identified AND the dataset is tested, report the accuracy value for the composited data.
- If a composited dataset is not tested, report the accuracy value for the least accurate dataset component.[197]

Similarly, Goodchild's and Gopal's 1989 work, *Accuracy of Spatial Databases*,[198] suggest that the cumulative effect of positional errors in several thematic layers (such as in a GIS overlay) is difficult to ascertain and may require multiple models for error.[199] This view is consistent with the suggested approaches in NSSDA mentioned above.

For CGD, a major source of positional error is due to the method for positioning features, which is to say the method that an end-user or contributor uses to establish the location of an object. Location information is typically entered using a variety of methods, including GPS coordinates (via surveying or recreational devices), Internet Provider (IP) geolocation, digitizing on a map or imagery, place names, or ZIP codes.

A thorough analysis of feature positioning methods in CGD was done by Brandon Shore (2012), who profiled the most common feature geometries and positioning methods in CGD applications (Figure 34). He determined that the most features were located from maps or imagery as digitized points, such as the creation of a placemark in Google Earth (Figure 35), followed by georeferencing by place name, georefencing by ZIP code, digitizing lines, and digitizing polygons. Point features in CGD may be collected in other ways as well, including GPS and IP geolocation. Each of these feature-positioning methods will be addressed briefly, with a discussion of typical accuracies and positioning characteristics.

---

[197] Authority, "Geospatial Positioning Accuracy Standards Part 1." Section 3.2.3

[198] Goodchild and Gopal, *The Accuracy of spatial databases*.

[199] Ibid., p. 33.

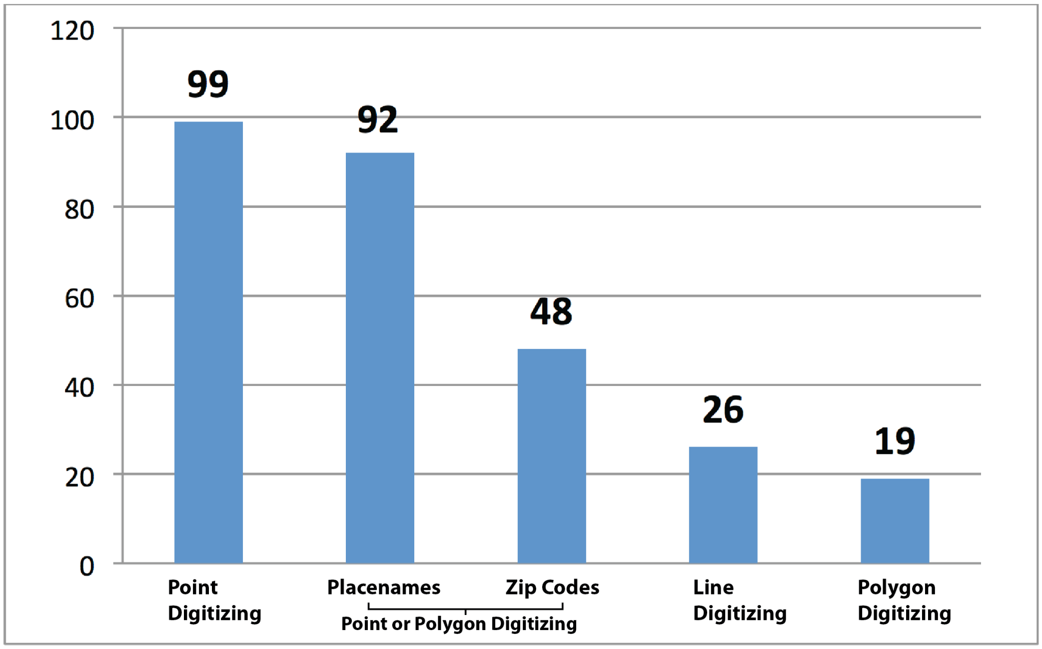


Figure 34. Feature Positioning in CGD[200]

Feature positioning using GPS

In May 2000, President Bill Clinton signed an executive order ending the longstanding practice of degrading the Global Positioning Signal available to consumers and businesses. The end of selective availability (as it was termed) led to enormous growth in the market for location-aware mobile devices, and therefore, CGD. Positional accuracy for civilian GPS systems went from +/- 150 meters to +/- 10 meters, overnight. Currently, the positional accuracy for standard civilian GPS devices, including those embedded in smartphones and personal digital assistants (PDAs), is thought to be around 10m.[201] The positional accuracy of GPS used in surveying applications is thought to be less than 1 m, and with post-processing corrections, as low as 2cm. [202] For CGD positioned with GPS, the positional accuracy associated with the standard civilian GPS applications (10m) is a very good proxy for positional error of the feature positioned with GPS. In CGD applications where transportation networks are generated and quality checked using GPS, multiple sources of position are typically collected

[200] Brandon M. Shore, "VGI Research Review" (presented at the AGC-VGI Research Review Meeting, Dr. Matt Rice, chair, George Mason University, Fairfax, VA, April 27, 2012).

[201] Paul A. Longley et al., *Geographic Information Systems and Science*, 3rd ed. (Hoboken, New Jersey: John WIley & Sons, 2011).

[202] Ibid.; "GPS Accuracy and Limitations," *Earth Measurement Consulting*, n.d., http://earthmeasurement.com/GPS_accuracy.html.

and compared to each other, with an end result for positional error much better than 10m. For features positioned using GPS, accuracy is not perfect, but is significantly better than many authoritative standards for positional error, such as the NMAS.

Features positioning using IP geolocation

All traffic on the Internet is routed with an addressing system developed in the early 1970s by Vint Cerf and Bob Kahn. The system uses an addressing system with 32-bit (four-byte) addresses, and more recently, 128-bit addresses. These addresses refer to locations on a network, and because the network is embedded in physical space, the addresses have some geographical association. For instance, all addresses in the block 129.174.XXX.XXX are assigned to the George Mason University campus in Fairfax, Virginia, and could be geolocated to the center of the campus. Other IP addresses, such as those originating from an Internet Service Provider (ISP), may be geolocated to the ISP's nearest network services center, which could be 10-15 miles away. IP geolocation is reasonably described to be accurate to 'city-level',[203] and consistent with the example of the George Mason University campus network, this is equivalent to anywhere from +/- 1.0 mile to +/- 10 miles. Several business offer IP geolocation services, claiming to have the most accurate and reliable geolocation services on the Internet, though there are no concrete positional accuracy statistics to enhance this claim.[204] Stefanidis et al. (2012) provide some examples of useful applications of IP geolocated CGD.[205]

Feature positioning using digitized points, lines, and polygons

Positioning features using digitized points, lines, and polygons is a very common technique, according to Shore (2012), with digitized points being the most common method, by far. Common web-based mapping programs, such as Google Maps and Google Earth, allow users to quickly position features using a single point digitized with a mouse click or fingertip on the top of a base map. Positioning of features using polygons and lines is done in similar fashion. The use of a push pin icon for the placemark

---

[203] Ian Devlin, "Finding Your Position with Geolocation," Blog, *HTML5 Doctor*, June 14, 2011, http://html5doctor.com/finding-your-position-with-geolocation/.

[204] For instance, "NetAcuity and NetAcuity Edge IP Location Technology," *Digital Element*, n.d., http://www.digitalelement.com/our_technology/our_technology.html.

[205] Stefanidis, Crooks, and Radzikowski, "Harvesting Ambient Geospatial Information from Social Media Feeds."

clearly indicates the marking of a single *x,y* point, even for an object such as Ayers Rock (Uluru) with a very large 2-dimensional footprint (Figure 35). Clearly, the single *x,y* positioning for a 2-dimensional object results in some imprecision in specifying position.

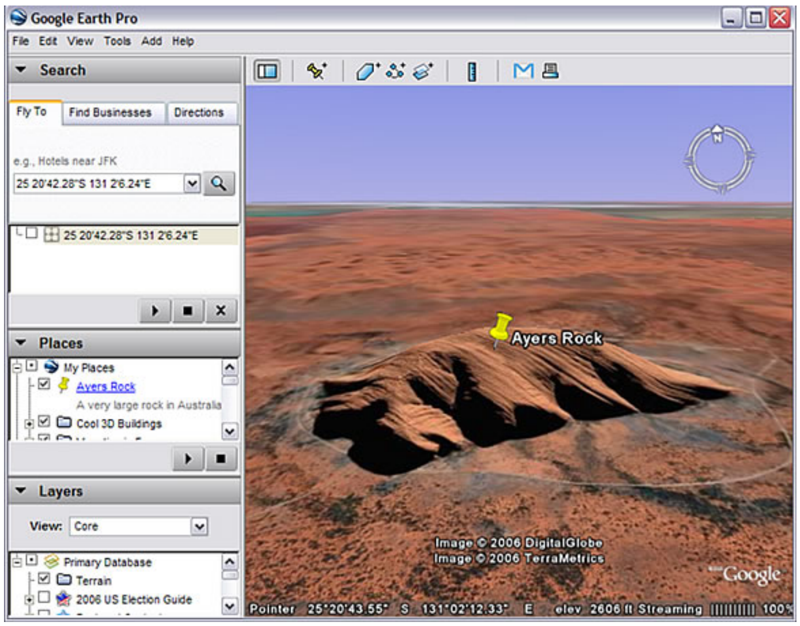


Figure 35. Positioning with a placemark: Ayers Rock (Uluru) in Google Earth

For features digitized with a digitizing table and puck, textbooks often cite the positional accuracy of points, lines, and polygons derived with this technique as 1/50th of an inch at map scale, which is equivalent to NMAS. For features digitized on a 1:24,000 scale base map, this suggests that positional error is approximately 40 feet, or 12 meters. With reference to Figure 35, the positional accuracy of features digitized on top of a base map will be directly related to the zoom level and latitude of the base map.

Feature positioning using place names

Shore (2012) also notes that place names are a common source for positioning, which can also result in error and imprecision, due primarily to ambiguity in feature identification where several proximate features share the same place name. The most common way of representing place name locations in a place name database or gazetteer is with a single coordinate location. Because of this, a building, city, state, or country place name are all recorded as a single location. This can create large positional errors, as

a coordinate associated with any place name is assigned the place names coordinate. As an example, a photograph taken in the state of California and tagged with the place name 'California' would be assigned a coordinate of -119.7512643, 37.2502247, if geotagged using the location of California from the U.S. Geological Survey's Geographic Names Information System. This could be hundreds of miles from where the photograph was taken. [206]

Feature positioning using ZIP codes

ZIP codes used for mail delivery in the United States are an important part of feature positioning in CGD. The initial 5 digital ZIP code, introduced in 1963 provided addressing support for sections of US cities, and the extended ZIP+4 code format introduced in the 1980s provides addressing support for very small geographic areas, including individual buildings or collections of 5-6 individual houses. A common misunderstanding of ZIP codes is the belief that they are polygons, when in fact they are simply a coded attribute of a collection of mail delivery points, and polygon-based representations are not produced or supported by the US Postal Service. Kahn (2012) studied common errors in ZIP codes represented as polygons, demonstrating large positional errors and significant logical errors that impact common spatial analysis procedures based on ZIP codes as polygons. For features positioned with 5-digit ZIP codes, the positional errors are highly variable within the range 1-5 miles.[207]

A final observation on positional accuracy from Shore (2012), is that positioning of features is often done with respect to a particular reference scale and a particular base-map rather than on actual position. Positional errors in the base-map data will translate into positional errors of the CGD positioned with respect to the base map. Rice (1998) explored visualization-based methods for correcting this type of relative positional error due to differences in base-maps.[208] In Shore's study, 78% of the 87 surveyed CGD applications used Google Map base data, 9% used OSM base data, 3% used ESRI data, 3% used NavTeq data, 2% used Microsoft Bing base data, and 2% used Google Earth data, each with their own slightly different positional characteristics. These different positional characteristics are

---

[206] "BGN: Domestic Names," *USGS*, April 9, 2012, http://geonames.usgs.gov/domestic/.

[207] Tunaggina Khan, "Evaluating the Errors Associated with Zip Code Polygon When Employed for Spatial Analysis" (MS Thesis, George Mason University, 2012).

[208] Matthew T. Rice, "A Visualization-based Method for Correcting Relative Positional Error Between Topographic Bases" (MS Thesis, Los Alamos National Lab and Brigham Young University, 1998).

passed along directly through the positions of CGD features. Because each base map can be displayed at a variety of different scales (Figure 33), the influence of the base map reference scale is thought to be a large contributor to positional error, more so than the individual differences between base maps.

Attribute Accuracy

Attributes, in a geospatial sense, are the non-spatial data linked to a location. Attributes describe the characteristics of a geospatial feature and can include anything from measureable characteristics, like length, width, temperature or wind speed, to descriptive characteristics, like ownership or land cover.

In the atomic model of geographic information discussed in Goodchild (2008) and Longley et al. (2011), attributes are the third element in a triple of {location, time, and attribute}. Location and attribute are often linked together with a relational database structure. Attribute accuracy, therefore, deals with the problems in correctly identifying the attributes associated with a location, as well as the incorrect assignment of numerical or text-based values associated with an identified attribute.

Longley et al.[209] discuss many of the problems related to ambiguity in definition and specification of attributes, as does Mark et al. (2007) who present a compelling study of the difficulty in finding common definitions for physiographic features, when translating between Aboriginal languages and English.[210] Terms for water bodies, in particular, differ significantly between English and Yindjibarndi, an Aboriginal language and ethnic group in Western Australia. The presence of subterranean water in otherwise dry stream channels and periodic surfacing or underground water sources is a part of Yindjibarndi landscape description, but not generally a part of English descriptions of the same features, reflecting the more direct association of the Yindjibarndi with the natural landscape. Another related aboriginal group has distinct terms for large pools, shallow pools, and transient pools formed by heavy rainwater, as well as an assortment of feature names for claypans, rock reservoirs, and sandy creek beds. English

---

[209] Longley et al., *Geographic Information Systems and Science*.

[210] Ibid.

attribute labels for the same features are much more limited, reflecting differences in definitions, identification, and use of landscape features.[211]

Errors due to misclassification and incorrect attribute values are common in CGD. If an attribute specification is available, this problem may be due to the contributor's inability to correctly assign the appropriate attribute. In some cases, assignment of an appropriate attribute value may be subject to interpretation, where even experts might disagree. In other cases, it may be due to a lack of expertise on the part of the contributor, who may lack the technical background required to understand and assign an appropriate value. For example, in Geo-Wiki.Org, which uses crowdsourcing to improve global landcover, the user must distinguish between categories from three different global land cover products, understanding such terms as "Mixed Forests," "Closed-Open mixed broadleaved forest," and "Tree Cover, needle-leaved, evergreen."[212] This requires specific, subject matter expertise.

Another issue arises in CGD systems, where users are allowed to tag or attribute features using their own terminology, a practice allowed in OSM. This can lead to inconsistencies, where different terms may refer to similar features, e.g., highway, motorway, freeway, autobahn; or where the same term may refer to different kinds features, e.g. the word 'village' may have different meanings in different contexts.[213]

OSM accommodates this difficult issue by providing guidance and examples, through their wiki, for the terms used to label features. While this falls short of a full attribute specification, it does bring order to the attribution process.

Logical Consistency

Logical consistency refers to the use of tests for validity of a feature, and is an important quality aspect for CGD. Tests for logical consistency include such things as checking for undershoots and overshoots (common topo-

---

[211] David M. Mark, Andrew G. Turk, and David Stea, "Progress on Yindjibarndi Ethnophysiography," in *Spatial Information Theory*, ed. Stephan Winter et al., vol. 4736, Lecture Notes in Computer Science (Berlin, Heidelberg: Springer Berlin Heidelberg, 2007), 1–19, http://www.springerlink.com/index/10.1007/978-3-540-74788-8_1.

[212] "The Geo-Wiki Project."

[213] Allan Brimicombe, *GIS, environmental modelling and engineering* (London; New York: Taylor & Francis, 2003).

logical errors when data has been digitized from a map), inspection for and removal of small sliver polygon (also a product of digitizing), and inspection for out-of-range data values.

There have been efforts to automate tests for logical consistency in transportation networks, as articulated by Goodchild,[214] who described such tests for intersections between secondary roads and freeways (Figure 36). In the United States interstate highway system, these intersections take the form of on-ramps, off-ramps, and cloverleaf interchanges, each with a specific geometry related to the expected traffic speed.

According to Goodchild, systems are being design to identify valid and invalid geometries, as part of a rules-based system for quality assessment in CGD. In Figure 36, a rule set could be developed to tag intersecting angles that are outside the valid range (right side). Importantly, these rules and tests for logical consistency can be developed from existing knowledge and existing datasets, which would be mined for significant patterns, relationships, and co-relationships.

Mooney and Corcoran[215] analyzed heavily-edited features in OSM data for the United Kingdom and Ireland, and noted that for the features that had been edited at least 15 times, 8% had invalid geometries which were never corrected.[216]

Girres and Touya[217] implemented specific topological tests for the presence of crossroads in the French OSM dataset, and determined that 5% of the crossroads had invalid topology.[218]

For CGD systems that are open (meaning no tests for validity are performed while features are being edited) there is no way to easily correct for errors in logical consistency. With OSM, there are multiple resources for

---

[214] Michael F. Goodchild, "Commentary on Future Trends and Research Objectives: Assessing the Value of Neo-Geographic Information" (presented at the AGC-VGI Research Review Meeting, Dr. Matt Rice, chair, George Mason University, Fairfax, VA, April 27, 2012).

[215] Peter Mooney and Padraig Corcoran, "Using OSM for LBS – An Analysis of Changes to Attributes of Spatial Objects," in *Advances in Location-Based Services*, ed. Georg Gartner and Felix Ortag, Lecture Notes in Geoinformation and Cartography (Springer Berlin Heidelberg, 2012), 165–179, http://dx.doi.org/10.1007/978-3-642-24198-7_11.

[216] Peter Mooney and Padraig Corcoran, "Characteristics of Heavily Edited Objects in OpenStreetMap," *Future Internet* 4, no. 1 (2012): 285–305.

[217] Girres and Touya, "Quality Assessment of the French OpenStreetMap Dataset."

[218] Ibid.

identifying and locating errors, like self-intersecting lines. The tools and a description of their functionality can be found on the OSM wiki.[219] However, no attempt is made to automatically correct these errors at the time of the data entry.

In the future, CGD projects may incorporate automated rules-based editing procedures that test for validity while the editing is being performed. This would improve data quality, but more importantly, would allow for user-feedback and training that would result in higher quality contributions.

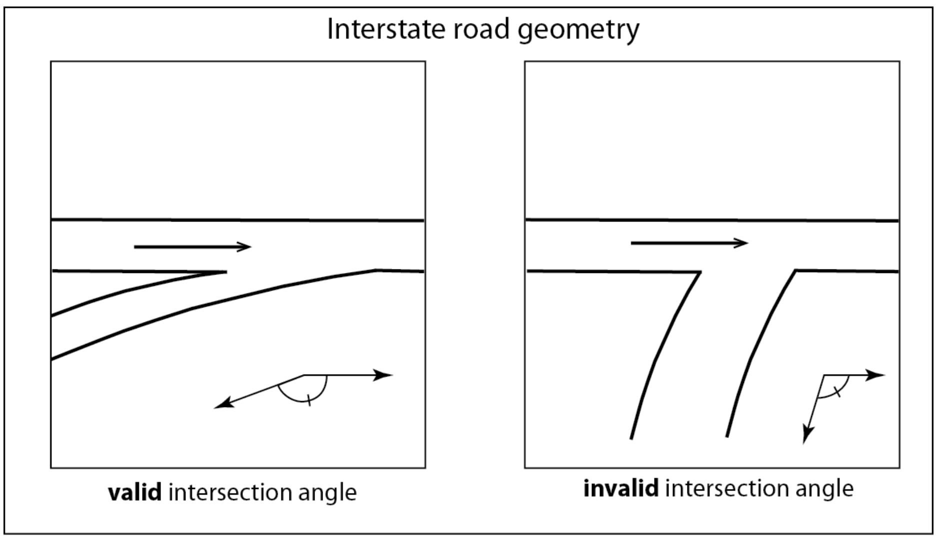


Figure 36. Future tests for logical consistency in CGD could automatically identify valid & invalid road geometries, based on a comprehensive rule set, such an acceptable intersection angles. For a US Interstate Highway, the angle on the right is too close to perpendicular

Completeness

Completeness refers to the comprehensiveness of included features in a dataset relative to the data's specification. The specification, in this context, describes the selection criteria and the amount of detail intended to be represented. As an example, if there were six public schools in an area, a dataset would be complete if all six public schools, as defined by the specification, were represented. It would be incomplete if only three schools were in the database.

---

[219] "Quality Assurance," *OpenStreetMap Wiki*, July 18, 2012, http://wiki.openstreetmap.org/wiki/Quality_assurance.

Veregin and Hunter[220] suggest that completeness can be measured in the spatial domain, temporal domain, or thematic domain. With regard to completeness, it is important to note that few large authoritative geospatial data collection projects are started without a target scale and level of detail in mind. Therefore, the data collections practices in these projects naturally result in the omission of certain geospatial details that are not relevant to the scale, level of detail, or general project specification.

If no prior database specification exists for purposes of assessing completeness (which is the case with many CGD projects), completeness cannot be assessed. Coverage, however, can be evaluated by comparing the dataset with an authoritative source at the same general scale and level of detail.

There is a distinction between completeness and coverage. Completeness is evaluated against a specification, while coverage is an assessment of the presence and density of features found in an area. Without a specification, it is not possible to determine when data collection is complete, but it is possible to assess the coverage.

Coverage in CGD has been measured and assessed by Haklay,[221] where the length of roads in OSM and the Meridian2 data were compared, showing that OSM had 69% of the coverage of the authoritative road dataset.

With CGD efforts, where volunteers determine which features to contribute, spatial coverage and completeness is a significant issue. Haklay's 2008 study found that OSM data has more coverage in affluent areas (76.6%) than in poor areas (46.1%). He also noted that there was more data in highly populated areas and less complete in rural errors. These differences reflect the preferences, biases, and local geographic expertise of CGD contributors, who are typically highly educated males.[222]

---

220 Howard Veregin and Gary Hunter, "Data Quality Measurement and Assessment," Educational resource, *The NCGIA Core Curriculum in GIScience*, 1998, http://www.ncgia.ucsb.edu/giscc/units/u100/u100_f.html.

221 Haklay, "How Good Is Volunteered Geographical Information?".

222 Mordechai (Muki) Haklay et al., "How Many Volunteers Does It Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information," *Cartographic Journal, The* 47, no. 4 (November 1, 2010): 315–322; "Po Ve Sham – Muki Haklay's Personal Blog," Blog, *OpenStreetMap*, March 5, 2012, http://povesham.wordpress.com/tag/openstreetmap/.

Temporal Quality

An advantage of CGD, as noted in the Chapter 2 discussion of production techniques, is the speed with which CGD can be produced. Zook,[223] Goodchild and Glennon,[224] and Ruitton-Allinieu[225] review the use of CGD production techniques and tools during urgent disaster response associated with California wildfires and the Haitian earthquake, contrasting the CGD approach with the much longer production techniques for authoritative data. Zook (2010), in particular, notes the value in combined or hybrid uses of CGD and authoritative data used during the earthquake. Many of the devices used for capturing CGD (smartphones, GPS, cameras, etc.) have the ability to capture time, and an acquisition time-date stamp is frequently embedded within the data.

Temporal quality in geospatial data is related to the accuracy of time measurements contained in the data, the dates when the phenomena being measured took place, and date and time when the data was recorded or entered in a database, the time periods for data validity, and most importantly (from the perspective of CGD), the update frequency for the dataset. The last item, update frequency, is important for CGD, due to the speed with which CGD can be collected.

Decades ago, authoritative production cycles could take years and typically ended with a map printed on a specific date. The production cycles for CGD are less discrete and are characterized by more frequent updates, with some data being available as soon as it is entered in the database. CGD becomes a very valuable tool for assessing rapidly unfolding events, as noted in Stefanidis et al.,[226] where social media were harvested to gain an understanding of geospatial footprints and associations of socio-political events.

---

[223] Matthew Zook, "Volunteered Geographic Information: Does It Have a Future?" (presented at the AAG Annual Meeting, New York, NY, 2012).

[224] Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

[225] Ruitton-Allinieu, "Crowdsourcing of Geoinformation."

[226] Stefanidis, Crooks, and Radzikowski, "Harvesting Ambient Geospatial Information from Social Media Feeds."

Non-Traditional and CGD Data Quality Concepts

A number of authors[227] discuss quality issues in CGD, starting with a context of traditional quality characteristics (as discussed in the previous section of this chapter). Each of the authors also cites quality characteristics and issues that are unique to CGD. Some of the unique characteristics involve community dynamics, crowd behavior, specifications, and rule-based triage of CGD. We also discuss malicious and mischievous content as an item for consideration in CGD data quality.

Malicious and Mischievous Content

An important consideration in quality assessments for geospatial data is the likelihood that the information being used is false. Authoritative data sources have very little worry about malicious content, as the production processes are controlled to a high degree. Rice[228] notes exceptions in the area of cartographic copyright traps, though the techniques usually have no bearing on usability. Generally, authoritative geospatial data is free of malicious content.

Wikipedia, perhaps the most prominent crowdsourced application in existence, has battled malicious content and vandalism for years, and has developed a number of analytical tools for detecting suspicious user editing activity. Although an automatic reaction to malicious and mischievous editing would be the imposition of user registration and accountability, the creators of Wikipedia recognize the benefit in maintaining an open system, which fosters higher levels of participation.

OSM, as the largest producer of CGD, has a sensible definition for what they consider to be vandalism. They define vandalism broadly to be "intentionally ignoring the consensus norms of the OpenStreetMap community," where users are expected to make "good accurate and well researched changes." They clarify that simple mistakes and editing errors

[227] Haklay, "How Good Is Volunteered Geographical Information?"; Girres and Touya, "Quality Assessment of the French OpenStreetMap Dataset"; Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content"; Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response"; David J. Coleman, Yola Georgiadou, and Jeff Labonte, "Volunteered Geographic Information: The Nature and Motivation of Produsers," *International Journal of Spatial Data Infrastructures Research* 4, no. 2009 (2009): 332–358.

[228] Matthew T. Rice, "Intellectual Property Control for Maps and Geographic Data" (Ph.D. Dissertation, University of California, 2005).

are not vandalism but can be fixed using the same tools that are available to fix vandalism.[229]

At the OSM project the tools for detecting and fixing vandalism include user profiling (white listing, user activity profiling[230]), difference and change detection algorithms,[231] and general data monitoring and analyzing code. The approach used by OSM is a combination of two general techniques for dealing with vandalism: automated checking and monitoring, and techniques based on Linus's Law (reviewed in this chapter), which suggests that errors can be corrected by the crowd and that the crowd will converge on the truth.

Examples of vandalism in OSM include text encoded with GPS tracks (Figure 37) and fake towns (Figure 38). Hidden content and artificial features of this type are well known within traditional cartographic works and have been extensively profiled by Monmonier[232] and Rice.[233] They are typically very minor features that have no impact on the usability of the geospatial data,[234] and would not be considered to be vandalism or malicious content. Generally, vandalism is not a problem in authoritative geospatial data production, and is nearly always associated with crowdsourced geospatial data production.

---

[229] "Vandalism," *OpenStreetMap Wiki*, July 10, 2012, http://wiki.openstreetmap.org/wiki/Vandalism.

[230] "UserActivity," *OpenStreetMap Wiki*, August 24, 2011, http://wiki.openstreetmap.org/wiki/UserActivity.

[231] "Osmdiff," *OpenStreetMap Wiki*, June 30, 2011, http://wiki.openstreetmap.org/wiki/Osmdiff.

[232] Mark Monmonier, *How to lie with maps*, 2nd ed. (Chicago, Illinois: University of Chicago Press, 1996).

[233] Rice, "Intellectual Property Control for Maps and Geographic Data."

[234] Ibid.

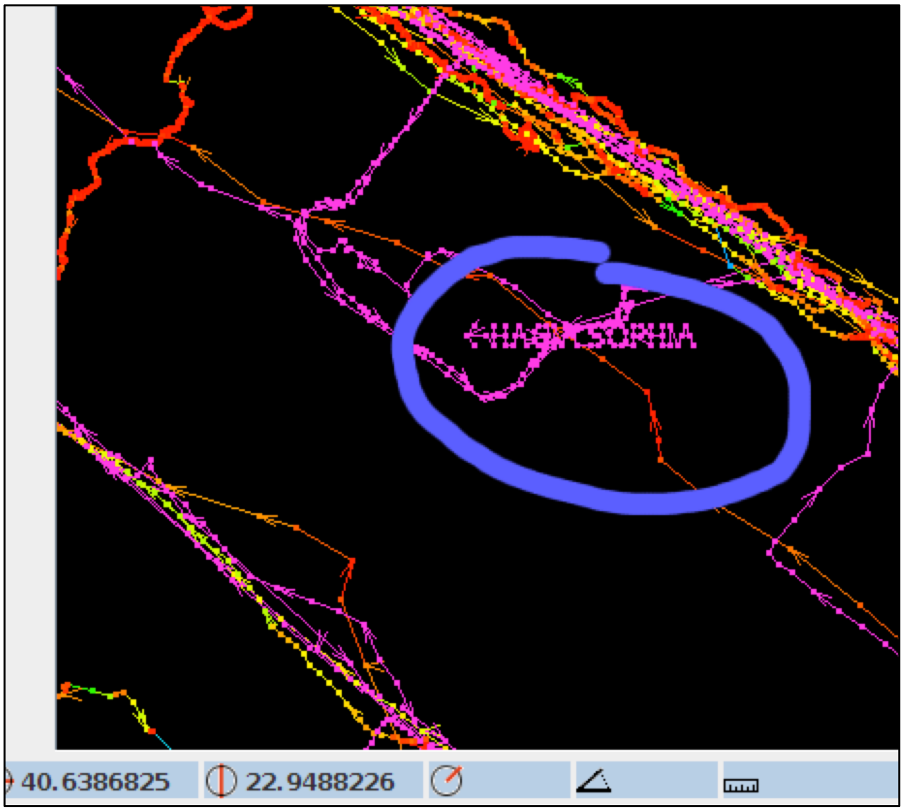


Figure 37. Text-based graffiti vandalism, encoded as GPS tracks, as seen in the OpenStreetMap Editor. GPS tracks read "HAGIA SOPHIA"[235]

Figure 38. Vandalism in the form of a fake town: West Harrisburg, Illinois[236]

235 *Gpx_graffiti_vandalism*, PNG Image, 1898 × 1130 pixels, n.d., http://wiki.openstreetmap.org/w/images/b/b7/Gpx_graffiti_vandalism.png.

A separate concern related to the use of smartphones is the integrity of the information collected, including location. Social media participants use the location and time information collected by smartphones to indicate their location, by content providers to customize services, and by businesses to facilitate transactions. As noted in the profile of Foursquare in Chapter 3, location information submitted by end-users can be intentionally misrepresented, resulting in problems and subsequent efforts to catch 'location cheaters'.

In cases involving business transactions, uncertainty about location and the integrity of the time-space data collected by the phone can have serious consequences, particularly for transactions where jurisdiction and local administrative issues are pertinent.

Lenders et al.[237] have developed a framework for a secure localization and certification service that can be used in social media and business transactions to increase trust in authenticity of content from mobile devices. Beach et al.[238] provide another approach that can be used to protect the end-user of the smartphone from unwanted invasions of privacy and breaches of security by mobile applications. Other approaches to preserve the integrity of smartphone time-location data will certainly emerge.

It is unlikely that many of the problems of malicious data associated with mobile device security and data integrity will be solved soon, but it is more than likely that CGD applications will continue to use smartphones and their sensors to gather and transmit geospatial data.

A final concern with malicious and mischievous data is the inclusion of profane or obscene content. In some instances, the use of offensive lan-

---

[236] *West_Harrisburg*, JPEG Image, 1716 × 954 pixels, n.d., http://wiki.openstreetmap.org/w/images/a/a2/West_Harrisburg.jpg.

[237] Vincent Lenders et al., "Location-based Trust for Mobile User-generated Content: Applications, Challenges and Implementations," in *Proceedings of the 9th Workshop on Mobile Computing Systems and Applications*, 2008, 60–64, http://dl.acm.org/citation.cfm?id=1411775.

[238] Aaron Beach, Mike Gartrell, and Richard Han, "Solutions to Security and Privacy Issues in Mobile Social Networking", vol. 4 (presented at the International Conference on Computational Science and Engineering, 2009. CSE'09., Vancouver, Canada, 2009), 1036–1042, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5283078.

guage is intended to deface a product. OpenStreetMap has adopted policies to rapidly remove inappropriate content, noting that it "might bring the project into disrepute."[239] In other instances, particularly with social media, profane and obscene language is a natural element of the conversations. A study of Yahoo! Buzz comments found that 9.4% contained profanity. [240]

Government agencies are particularly sensitive to broadcasting social media content that might be offensive, as illustrated by the Department of the Interior (DOI) policy that clearly states that they monitor social media contributions and reserve the right to delete "violent, vulgar, obscene, profane, hateful, or racist comments."[241]

Automated filters, often based on lists of banned words, may assist in the process of monitoring content. This approach, however, has limitations due to "misspellings (both intentional and not), the context-specific nature of profanity, and quickly shifting systems of discourse that make it hard to maintain thorough and accurate lists."[242] Because of these issues, manual review may be necessary to fully police content.

The ability to control malicious content and vandalism is essential to the success of any project using CGD. Sufficient resources, including tools and manpower, must be identified and allocated in order to insure the integrity of the project.

Balancing Adherence to Specifications with User Participation

Girres and Touya[243] suggest that one of the major reasons that OSM and other CGD datasets have quality problems is a lack of formal specification in the creation of the dataset:

> The evaluation of the different aspects of OSM data quality . . . reveals the key role of specifications to ensure quality, as several error types come from a lack

[239] "Vandalism."

[240] S.O. Sood, J. Antin, and E.F. Churchill, "Profanity Use in Online Communities" (2012), http://research.yahoo.net/files/profanity_chi.pdf.

[241] W. E. Ricker, *Computation and Interpretation of Biological Statistics of Fish Populations*, vol. 191 (Fisheries and Marine Service, 1975).

[242] Ibid.

[243] Girres and Touya, "Quality Assessment of the French OpenStreetMap Dataset."

> of, or fuzzy, specifications . . . In OSM, the specifications are rich and complex but informal, instead of being recorded in written formal and well-accepted specifications. A contributor is advised to follow the specification but does not have to.[244]

Citing Coleman,[245] Girres and Touya note that the majority of CGD contributors are not subject domain experts, but are “occasional contributors and mostly *interested amateurs* who could be afraid of strict specifications for contributions.” They go on to note:

> The success of VGI lies in the simplicity of contributions, and many debates in the OSM contributor community show that this should not be too restrained, even to improve quality. We believe that the improvement of OSM data quality requires finding the ideal balance between specifications and contribution freedom.[246]

The success of CGD is in the simplicity of the contributions, and enforcing adherence to strict specifications and quality guidelines will result in reduced contributions. Ultimately, there must be a balance between required specifications, quality control, and contribution freedom.

Linus’s Law

Goodchild,[247] notes other important CGD-related data quality issues that are significant and will be addressed. First, a major argument in favor of crowdsourcing any content is linked to Linus’s Law. The law states that if enough eyes (participants, contributors) review a problem, the remedy or solution will be obvious to someone, who will quickly make the necessary correction.[248]

---

[244] Ibid., p.457

[245] Coleman, Georgiadou, and Labonte, “Volunteered Geographic Information.”

[246] Girres and Touya, “Quality Assessment of the French OpenStreetMap Dataset.” p.457

[247] Goodchild, “Assertion and Authority: The Science of User-Generated Geographic Content”; Michael F. Goodchild, *Assessing the Value of Neo-Geographic Information: Report on Project Review, 4/27/2012*, Unpublished Project Review Report (Fairfax, VA: George Mason University, April 27, 2012).

[248] Eric S. Raymond, *The Cathedral and the Bazaar*, ed. Tim O’Reilly, 1st ed. (Sebastopol, CA, USA: O’Reilly & Associates, Inc., 1999).

Goodchild and Glennon,[249] note that this approach works quite well, as evidenced through comparisons of crowdsourced materials such as Wikipedia and objective comparisons of article quality with traditional encyclopedia. Given the large number of people editing and contributing to Wikipedia, at least one person (in line with Linus's Law) will have the topical expertise to help the article converge toward 'truth'. Goodchild[250] suggests that this is a commonly used argument in favor of CGD from a quality standpoint.

Because CGD contributors may be widely distributed, the topical expertise, cited as a contributing factor in Wikipedia's success, is replaced with local geographic expertise, which everyone has as a product of their experience in life and activity spaces.[251] This improves the quality of the contributions, particularly for data in areas where the contributors are most knowledgeable, which may include areas where a contributor currently lives, has lived in the past or has visited. For some types of data, like geotagged photos, the individual must be at the specific location to capture the data.

Several CGD applications that focus on local geographic expertise, such as GasBuddy, NavTeq Map Reporter, FourSquare, GrassRoots Mapping, and WikiMapia, are profiled in Chapter 3.

This does raise an issue in regard to Linus's Law. For Linus's Law to work well in cases where local geographic expertise is required, a CGD project would need to have a body of contributors who are broadly and uniformly distributed throughout the study area to help the CGD resource converge toward geographic 'truth'. An analysis of CGD contributions, such as that by Haklay,[252] shows completeness problems, with bias for areas of affluence and bias against poor areas. This issue becomes more significant when time is a critical component of the CGD, as found in applications like Waze, which monitor current traffic. In this case, a critical mass in both space and time is needed to provide enough relevant data.

There are few examples, however, of projects big enough, or with a large and diverse enough contributor community, to convincingly provide con-

[249] Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

[250] Goodchild, *Assessing the Value of Neo-Geographic Information: Report on Project Review, 4/27/2012*.

[251] Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content."

[252] Haklay, "How Good Is Volunteered Geographical Information?".

vergence toward 'truth' for large geographical areas where local geographic expertise is required. OSM is the exception here. Conceivably, if a CGD project had a very narrow data specification and a very small geographical area, a large enough contributor pool could be found or recruited that would function in the way imagined by Linus's Law.

Not all successful CGD projects require local expertise. Projects covering a large area, but requiring little local expertise, like Galaxy Zoo, can also succeed. In Galaxy Zoo, contributors simply describe patterns using commonly understood terms. Transcription and imagery-based search problems are other common geospatial tasks that do not require local expertise.

CGD projects rely on two primary strategies for assessing quality: serial review and multiple collects. With serial review, contributors edit the work of others, continually adding to the content and improving the quality of the data. This is the approach taken by OSM, Google Map Maker, and Wikimapia. With multiple collects, a number of different contributors work on the same task and their results are compared. This technique has been used effectively in Old Weather, Galaxy Zoo, and the Field Expedition: Mongolia.

All projects require review and editing. From a quality perspective, the key issue for crowdsourced data quality is to evaluate the review process to insure that sufficient 'eyeballs' are available to identify and correct errors.

Hierarchal Structures for Quality Assurance

A quality assurance method built into many crowdsourcing projects, including Wikipedia, is a social or community-based mechanism, where a hierarchy of moderators and gate-keepers are used to check contributions from lower-level participants.

In some cases, these moderators have specific domains and areas of expertise, and are promoted based on their track records and level of familiarity with project guidelines, specifications, and protocols. They are called upon the solve disputes and make judgments on items where consensus is not clear. In other cases, like Google MapMaker, these gate-keepers may be company employees evaluating data to corporate standards.

Goodchild[253] suggests that a geographical version of this hierarchical control mechanism could be constructed, but that like any bureaucratic structure, it would have the disadvantage of slowing the release of data, particularly in disaster response scenarios, where an immediate response is needed.[254]

Rules-Based Triage of CGD

A third approach for quality assurance in CGD projects, suggested by Goodchild,[255] is an automated triage approach, where contributions are assessed against a rule-base, which contains a distillation of significant patters, relationships, and co-relationships. This rule-base could be constructed from previously quality-checked authoritative data and would be used during data production to automatically flag errors, with the added benefit of providing guidance to less experience contributors.

Simpler versions of these rule-bases are already used, to some extent, in geospatial software. For instance, one procedure that is frequently done automatically before creating a digital elevation model is the filling of 'sinks' or topographical regions without any external drainage. Natural topographic sinks are rare, and are more commonly the result of data quality problems. Similar triage could be easily done with road networks, to ensure topology and to correct for errors such as that shown in Figure 36 (right side) where the entrance to an Interstate freeway is incorrect.

Documenting Geospatial Data Quality - Metadata

Metadata, or "data about data," can be thought of as a summary of the content and context, of a dataset. Metadata is typically created by the producer of authoritative datasets, and is an important means for communicating and conveying basic information about the data, as well as quality information.

Metadata can be collected for observations, datasets, or collections of datasets. While authoritative geospatial datasets frequently have dataset and collection-level metadata meeting formal national or international stand-

---

[253] Goodchild, *Assessing the Value of Neo-Geographic Information: Report on Project Review, 4/27/2012*.

[254] Ibid.

[255] Ibid.

ards, dataset level metadata is commonly lacking in CGD, although it may be possible to extract metadata about individual contributions.

Some observation metadata is generated automatically, for instance, by digital cameras that have a capability of attaching geotags, time stamps, camera settings and other information in the headers of digital image files.

Dataset metadata is recorded by the individual or group collecting the data. For authoritative data, this is often done to formal geospatial metadata standards such as the Content Standard for Digital Geospatial Metadata (CSDGM),[256] the International Standards Organization suite of metadata standards,[257] or the Dublin Core Metadata standards.[258]

The formal metadata standards tend to be complex and difficult to understand, resulting in metadata that is frequently incomplete or missing. CGD efforts generally do not collect formal dataset level metadata.[259] As noted in Chapter 2, even authoritative data generated by well-funded government agencies can be faulty, lacking basic attribute data and metadata.

As an alternative to formal metadata, others have also suggested implementing informal metadata in the form of folksonomies, which are tags or descriptors attributed to particular items.[260] A key example of a successful implementation of folksonomies linked to individual contributions is Flickr. Bishr and Kuhn believe that improved and better-designed folksonomies will lead to better knowledge extraction as well as more being gleaned from the tagging and querying process.

They point to a website, Tidepool,[261] as an example of how to set up such a process. Tidepool provides four possible tags for users to fill in which in-

---

[256] "Geospatial Metadata Standards — The Content Standard for Digital Geospatial Metadata (CSDGM)," *Federal Geographic Data Committee*, April 25, 2012, http://www.fgdc.gov/metadata/geospatial-metadata-standards#csdgm.

[257] "Geospatial Metadata Standards — FGDC Endorsed ISO Metadata Standards."

[258] "Dublin Core Metadata Element Set, Version 1.1," *Dublin Core Metadata Initiative*, June 14, 2012, http://dublincore.org/documents/dces/.

[259] Rodolphe Devillers et al., "Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities," *Transactions in GIS* 14, no. 4 (August 2010): 387–400.

[260] Mohamed Bishr and Werner Kuhn, "Geospatial Information Bottom-Up: A Matter of Trust and Semantics," in *The European Information Society - Leading the Way with Geo-information*, ed. Sara I. Fabrikant and Monica Wachowicz, Lecture Notes in Geoinformation and Cartography (Springer-Verlag Berlin Heidelberg, 2007), 365–387, http://dx.doi.org/10.1007/978-3-540-72385-1_22.

[261] "Storymill - Tidepool," *Storymill*, 2012, http://storymill.com/tidepool/.

clude, who, where, when, and what. A process such as this can easily be transitioned to geographic data with an added emphasis placed on the 'what' and 'when' categories. CGD repositories like ArcGIS Online and GeoCommons support dataset level tagging.

Efforts to encourage the production and preservation of metadata should be made. The proper use of metadata is a necessary precursor to interoperability and use of CGD with other datasets.

Summary

There are a number of traditional data quality measures that are well-developed and understood that can be applied to CGD, including lineage, positional accuracy, attribute accuracy, logical consistency, and completeness. A few of the more prominent measures have been discussed in this chapter. A few data quality ideas, more directly applicable to CGD, have also been presented, including Linus's Law, hierarchal structures for quality assurance, and rules-based triage of CGD. An important final aspect of quality is the use of metadata to record summaries of a dataset's contents and context.

Together, these various traditional and CGD-based quality ideas will shape the future development of quality assurance for CGD, particularly for hybrid geospatial data projects that contain a mixture of authoritative and crowdsourced data. The next chapter addresses the evaluation of CGD, and offers ideas and considerations to be used when considering the use of CGD.

5 Evaluating Crowdsourced Geospatial Data

Crowdsourced geospatial data (CGD) is a new, emerging phenomenon, presenting opportunities and risks. As mentioned in the previous chapter, there are many ways to assess and evaluate the quality of CGD, and the evaluation will ultimately help determine whether CGD can be a useful part of a specific geospatial project or enterprise. Evaluation is a key part of considering whether or not to use CGD and adopt CGD-based production methods. This chapter addresses several key topics for consideration in evaluating CGD for adoption.

Reviewing metadata is the most basic method to evaluate CGD data quality. If formal or informal metadata is not available, insufficient, or faulty, CGD can be evaluated using techniques and methods described in Chapter 4, "Quality and Crowdsourced Geospatial Data" and the remainder of this chapter.

Visualizing Uncertainty

For CGD where some measure of uncertainty is available, visualization can be a useful way to assess and evaluate the usefulness of CGD, and the techniques for visualizing this uncertainty are similar to techniques used with authoritative data. For authoritative data, there are hundreds of research papers on assessing and visualizing uncertainty. The National Center for Geographic Information and Analysis (NCGIA) conducted a thorough year-long investigation of techniques for visualizing the quality of spatial data.[262] Other groups, such as the International Cartographic Association (ICA), have sponsored a large number of publications on visualization in cartography and techniques for visualizing data quality.[263]

Paradis and Beard, affiliated with the NCGIA during the early 1990s, proposed a novel technique for visualizing and communicating spatial data quality to decision makers.[264] They present a data quality filter to organize

[262] "NCGIA Research Initiatives - Initiative 7: Visualizing the Quality of Spatial Information," *NCGIA*, 1993, http://www.ncgia.ucsb.edu/research/initiatives.html.

[263] "ICA Commission on GeoVisualization," *ICA ACI*, July 11, 2011, http://geoanalytics.net/ica/.

[264] P. Mooney, P. Corcoran, and A.C. Winstanley, "Towards Quality Metrics for Openstreetmap," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, 514–517.

and communicate uncertainty to a decision maker. This filter consists of user-entered values for uncertainty, and translates them directly to a visualization-based method for depicting the uncertainty. For instance, in a dataset where positional error varies and a threshold quality limit is established at 30 meters, data points can be displayed using variable density, with points having unacceptably high levels of error shown as empty circles, and points having acceptable positional errors shown using a solid color, as seen in the figure below (Figure 39). For CGD with any estimates of quality, this method of filtering could be very useful for visually tagging the CGD that fall below a quality threshold.

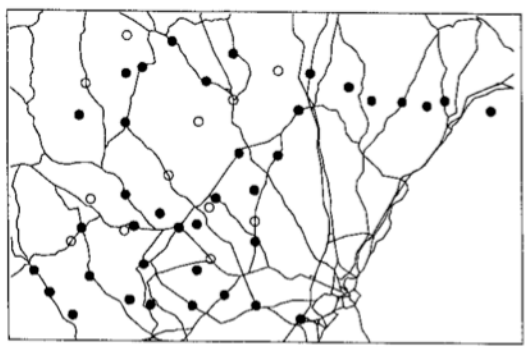


Figure 39. Data values not meeting a user-entered positional accuracy threshold of 30 meters are depicted with open circles[265]

MacEachren et al.[266] present a number of useful methods for representing the reliability in georeferenced health statistics data, including the use of side-by-side maps of data and associated uncertainty (Figure 40) and maps with data and uncertainty toggled on and off using the computer mouse (Figure 41).[267] MacEachren and others have done extensive user testing and evaluation to determine the best methods for displaying combinations of geospatial data and associated uncertainty or error. A good summary of this and similar research is available through Slocum et al.[268]

---

[265] Jeffrey Paradis and Kate Beard, "Visualization of Spatial Data Quality for the Decision Maker: A Data Quality Filter," *URISA Journal* 6, no. 2 (1994): 25–34.

[266] A. M. MacEachren, C. A. Brewer, and L. W. Pickle, "Visualizing Georeferenced Data: Representing Reliability of Health Statistics," *Environment and Planning A* 30 (1998): 1547–1562.

[267] Adrienne Gruver, "Concept Gallery," Educational resource, *Penn State - College of Earth and Mineral Sciences*, 2012, https://www.e-education.psu.edu/geog486/l8_p5.html.

[268] Terry A. Slocum et al., *Thematic cartography and geovisualization*, 3rd ed. (Indianapolis, Ind.; London: Prentice Hall; Pearson Education [distributor], 2009).

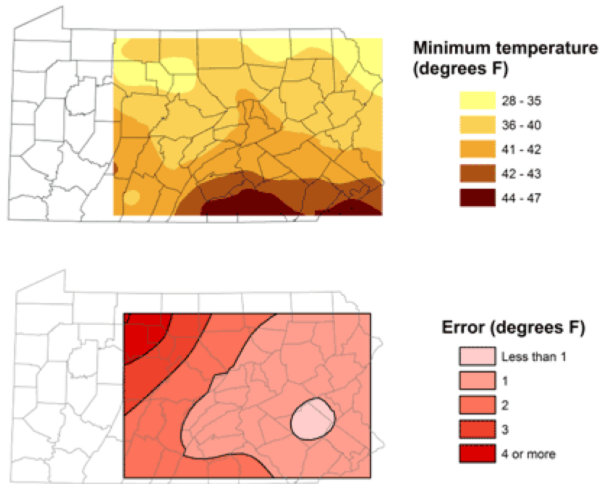


Figure 40. Maps of minimum temperature and error, shown side by side for visual comparison[269]

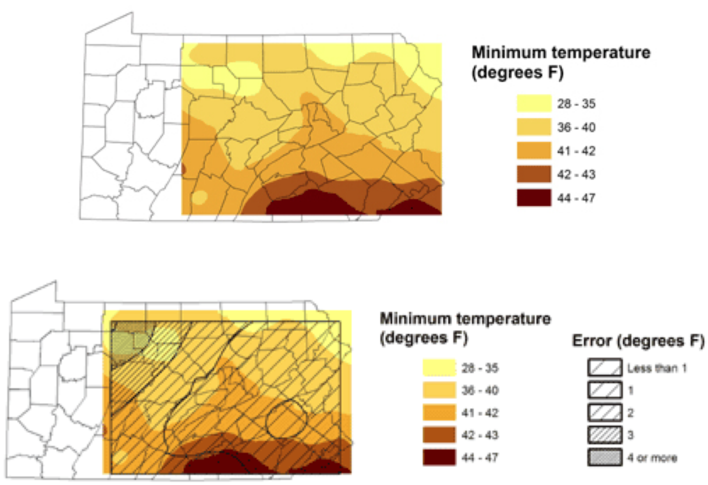


Figure 41. A single map of minimum with temperature and error layer alternately toggled off (top) and on (bottom) with the computer mouse[270]

Mooney et al.[271] present several methods for measuring, assessing, and visualizing uncertainty in the geometry of features and metadata in OSM. They use a simple overlay for visual comparison of shape (Figure 42),

---

[269] MacEachren, Brewer, and Pickle, "Visualizing Georeferenced Data."

[270] Ibid.

[271] Mooney, Corcoran, and Winstanley, "Towards Quality Metrics for Openstreetmap"; Mooney, Corcoran, and Winstanley, "A Study of Data Representation of Natural Features in Openstreetmap."

shape similarity statistics for comparing OSM and Ordnance Survey Ireland (OSI) data (Figure 43), and histograms of polygon vertex density for shape uncertainty comparisons (Figure 44). The combination of visual overlay and simple statistical shape comparisons is a useful way to assess quality and determine the usability of CGD for various applications.

Figure 42. Overlay of OSM data and orthoimagery of hydrologic features for visual comparison and assessment of uncertainty[272]

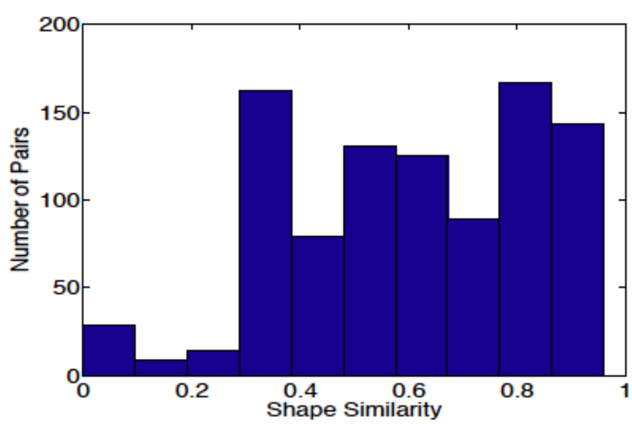


Figure 43. Histogram of shape similarity statistics for OSM and OSI data for quality assessment[273]

[272] P. Mooney et al., "Citizen Generated Spatial Data and Information: Risks and Opportunities," in *Proceedings of the 2nd International Conference on Network Engineering and Computer Science (ICNECS 2011)*, 2011.

[273] Ibid.

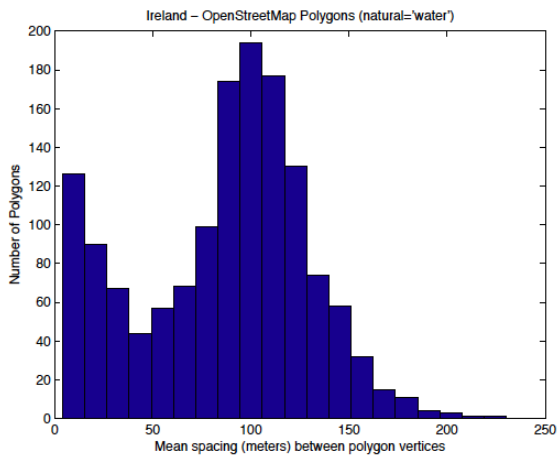


Figure 44. Histogram of polygon vertex density for OSM and OSI data for quality assessment[274]

Ruitton-Allinieu[275] presents a very useful and comprehensive summary of quality assessment in CGD, including many visualization-based methods for depicting uncertainty and quality issues in CGD. She uses examples from Haklay,[276] Girres et al.,[277] and other works to show how visualization based uncertainty estimates of CGD can be constructed.

Comparison to a Reference Resource

One useful and longstanding method for assessing the uncertainty and suitability of geospatial data is through comparisons to a reference source. The older map-based National Map Accuracy Standards and the more recent National Standard for Spatial Data Accuracy provide standardized methods for verifying and assessing positional accuracy.

Crowdsourced geospatial data users can perform the analyses themselves or review analyses prepared by experts. Several useful studies, cited in Chapter 2 and Chapter 4 of this report, compare the quality of crowdsourced geospatial with authoritative data. These studies[278] are summarized by Ruitton-Allinieu.[279]

---

[274] Mooney, Corcoran, and Winstanley, "A Study of Data Representation of Natural Features in Openstreetmap."

[275] Ruitton-Allinieu, "Crowdsourcing of Geoinformation."

[276] Haklay, "How Good Is Volunteered Geographical Information?".

[277] Girres and Touya, "Quality Assessment of the French OpenStreetMap Dataset."

[278] Haklay, "How Good Is Volunteered Geographical Information?"; Girres and Touya, "Quality Assessment of the French OpenStreetMap Dataset"; Mooney, Corcoran, and Winstanley, "A Study of Data Representation of Natural Features in Openstreetmap"; Zielstra and Zipf, "Quantitative Studies on the Data Quality of OpenStreetMap in Germany."

[279] Ruitton-Allinieu, "Crowdsourcing of Geoinformation."

Haklay's research assessment of OSM and Ordnance Survey data is well known, and an often-cited work on comparisons between CGD and reference data. For England, the OSM roads dataset had an average position error of approximately 6 meters when compared to the authoritative Ordnance Survey data. A comparison of OSM and Ordnance Survey datasets for motorways found that their positions overlapped 80% of the time. The OSM data, however, has less coverage than the Ordnance Survey data, with approximately 57% of all Ordnance Survey roads covered in the OSM dataset.

In a related study, Girres et al.[280] discovered that OSM road networks in France have relatively good topological consistency, with less than 5% of street intersections in error. In comparison, the authoritative source of the data for the same area had an error rate of less than 1%. They found, however, that OSM place names for lakes contained errors at least half the time in comparison to the authoritative, government naming.

The Christmas Bird Count, profiled in Chapter 3, has been compared to the authoritative US Fish & Wildlife Service's Breeding Bird Survey. Dunn et al.[281] summarize a variety of previous studies that have found correlations between changes in bird abundance noted in the Christmas Bird Count and the same characteristics measured in the Breeding Bird Survey. They also note strong correlations between the Christmas Bird Count and Project FeederWatch, a more standardized winter bird count.[282]

If no reference source exists, comparisons of CGD with aerial photos, satellite imagery, and textual sources may be useful.

User Experience

CGD may also be assessed by visualizing it or applying it directly in geospatial analysis. Visual assessment can often be used to quickly identify anomalies and erroneous data. In the example below, data errors associated with ships logs are evident where routes cross over land.

[280] Girres and Touya, "Quality Assessment of the French OpenStreetMap Dataset."

[281] Erica H. Dunn et al., "Enhancing the Scientific Value of the Christmas Bird Count," *The Auk* 122, no. 1 (2005): 338–346.

[282] Ibid.

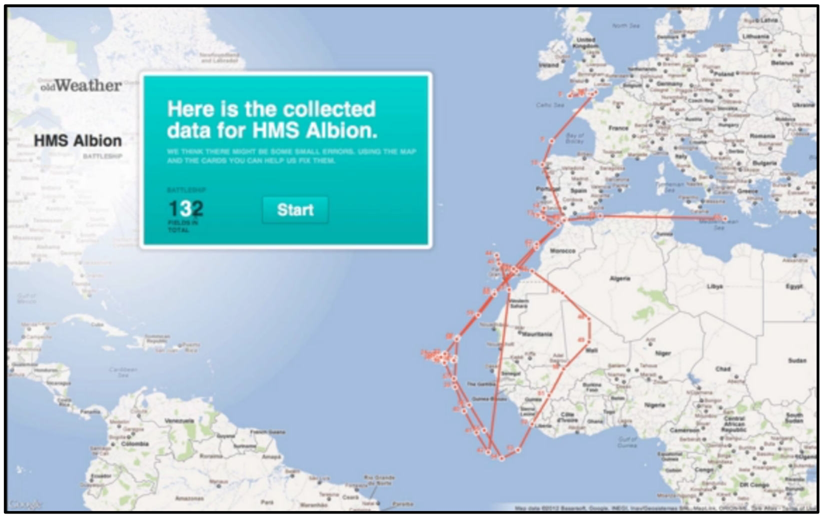


Figure 45. Visual assessment of Old Weather Data. It is possible to identify errors in ships log entries by visualizing the data. In this example, the entries located in western Africa are obvious errors[283]

Visual assessment can be accompanied by analytical evaluation. If reference data exists, analytical results using CGD can be compared with the results using reference data. This is especially useful for navigation and routing data, but can be applied to any analytical operation.

Where no reference data exists, results can be compared with expected results. While visual analysis is sometimes sufficient for determining the quality of a data source, testing is especially valuable for applications like routing, where topological errors that might be difficult to detect visually can produce unexpected paths (Figure 46).

[283] "Old Weather Review Interface," *Vimeo*, n.d., http://vimeo.com/39450854.

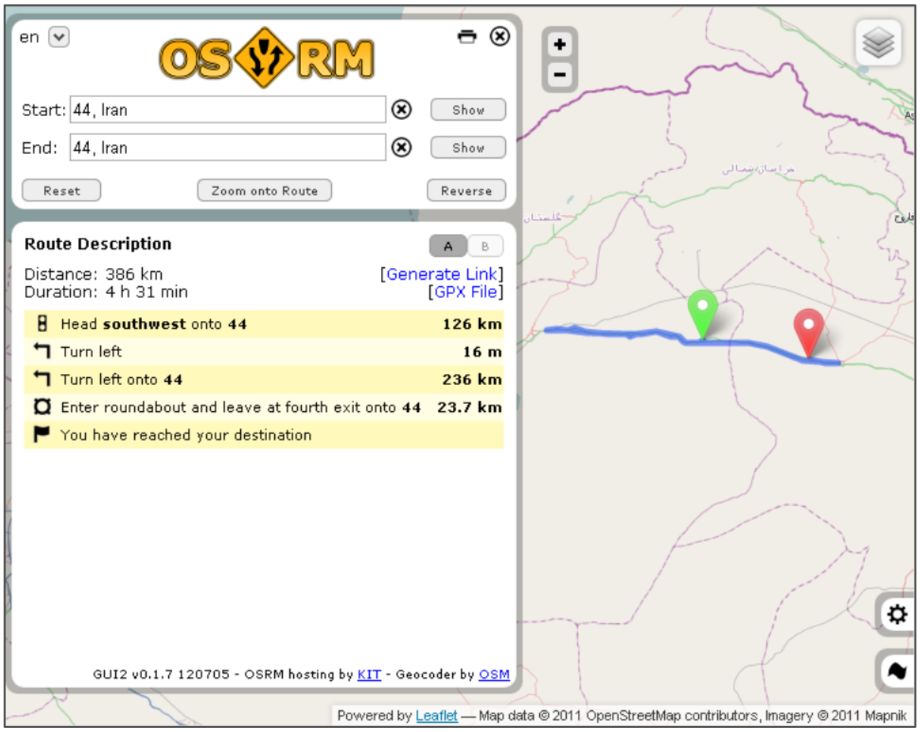


Figure 46. Unexpected origin-destination routing produced by topological errors in OSM base data, Open Source Routing Machine[284]

Expert and User Reviews

Expert and peer review is a fundamental part of science. The evaluation of a scientific work by peers is seen as a necessary step toward improvement and eventually, truth. In academia, other experts do peer evaluation. Expert CGD assessments often take the form of studies comparing CGD against authoritative data sources, as described in the previous section.

Alternatively, simply having an expert or an authoritative organization adopt a geospatial dataset is a form of review and endorsement. The rapid adoption of OSM software and data during the earthquake by disaster response group led the US Military's Southern Command (SOUTHCOM) to adopt OSM.[285] By having SOUTHCOM adopt CGD data during this crisis, the threshold for other Government organizations use of the OSM data was lowered.

---

[284] "OSRM Website," *OSRM*, 2011, http://map.project-osrm.org/.

[285] Zook et al., "Volunteered Geographic Information and Crowdsourcing Disaster Relief."

User ratings complement expert reviews and opinions. Unfortunately, the term is used, confusingly, to describe the assessments of a contributor as well as the assessment of the contributor's contributions. Both types of assessment are important, and in the case of meritocratic production systems, the two types of ratings are assumed to be highly correlated.

User ratings, in the sense of contributor evaluations, are most appropriate for CGD where the contributions can be directly linked to an individual, such as images contributed to Flickr or tweets on Twitter. In this case, higher contributor ratings are typically correlated with higher quality contributions, although any individual contribution may be of lower quality.

When data is produced collaboratively, like OSM, or as a result of comparing multiple contributions, like Old Weather, content rating systems are more appropriate. In these cases, assessment is more difficult as it is no longer possible to attribute the final data to a single individual.

It may be sufficient to evaluate the overall quality of the data set, but in some instances the quality information about individual features is required. Content rating can be done by peers or other users, as well as by automated analysis. Although content rating can be applied to individual elements, like sentences in an article or individual features in a database, it is typically applied at a higher level, such as an entire database.

Wikipedia's article rating system, introduced in 2011, allows users to assess an article's trustworthiness, objectivity, completeness, and writing quality. In addition to ratings by users, automated rating systems have been developed for Wikipedia to rate the quality by evaluating metadata, trustworthiness, author reputation, and revisions. Applying similar techniques to CGD is an area of future research.

Content rating by users is frequently done using a simple scale supported by comments. Some ratings are based on simple approval or disapproval, like thumbs up or thumbs down, while others are based on a numerical scale, such as ArcGIS Online's 5-star rating system. Amatriain et al.[286] develop a method to characterize user ratings variability within media rec-

[286] Xavier Amatriain, Josep M. Pujol, and Nuria Oliver, "I Like It... I Like It Not: Evaluating User Ratings Noise in Recommender Systems," in *User Modeling, Adaptation, and Personalization*, ed. Geert-Jan Houben et al., vol. 5535 (Berlin, Heidelberg: Springer Berlin Heidelberg, 2009), 247–258, http://www.springerlink.com/index/10.1007/978-3-642-02247-0_24.

ommendation systems, noting that users are sometimes inconsistent in giving feedback.[287]

Risk Management

In assessing whether CGD would be useful for a particular task, it is critical to determine what the possible impacts would be if the information were incorrect. For scientists, this line of reasoning is formalized in statistical hypothesis testing, where a Type 1 error represents the probability of rejecting a null hypothesis that is actually true, and a Type 2 error represents accepting a false null hypothesis. Scientists often construct hypothesis tests to minimize the chances of a Type 1 error (analogous in US criminal courts to convicting an innocent person). A Type 1 error is often viewed much less palatable than a Type 2 error (letting a guilty person go free).

Similarly with CGD, if the information contained in the geospatial data is incorrect, an assessment needs to be made about how serious a problem that would be. The risk presented by information being incorrect would be balanced against the benefit obtained by using the information.

For situations involving critical danger to personnel and resources, this balancing would be done in a very conservative manner, and with a deliberate process. First, the risks for using incorrect or erroneous CGD would be identified. Second, the risks would be analyzed and evaluated in the context of potential dangers, and the risks would be documented.

The scenarios described by Goodchild and Glennon[288] in the Santa Barbara, California wildfires represent real events and significant risks. As the wildfire moved through the Santa Barbara area and the neighborhoods evacuated, residents had to carefully weigh the risk of evacuation, with its extreme stress, discomfort, and dislocation, with the risk of staying in place (possible injury or death).

Clearly, the risk of staying in place during an advancing wildfire is a much more serious problem, akin to our Type 1 statistical error. During this wildfire event, crowdsourced maps showing detailed fire boundaries were available, and appeared to be updated much more frequently than the au-

---

[287] Ibid.

[288] Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

thoritative maps. The presence of the CGD-based maps allowed many residents to more carefully determine when to evacuate and weigh risks of staying in place.

Gervais et al.[289] formalize the reasoning discussed in this chapter by proposing and implementing a risk management concept in a spatial database and spatial online analytical processing (SOLAP) framework. Their approach is suggested to create a more formal responsibility relationship between developers (those that would create CGD and store it in a database) and users.[290] Any mechanism to add a sense of responsibility for potential risks to those creating CGD for others to use is a very good development.

Ultimately, every use of CGD in emergency scenarios or critical situations will need to be weighed carefully. There are many methods for assessing quality and using that assessment in a decision making process. Unlike traditional map data, which is assumed, often wrongly, to be correct in its entirety, CGD might be best characterized as an intelligence resource that is partial, incomplete, with risks and potential benefits. Depending on the situation and the use, it may be very useful, or may represent an unacceptable risk. The next chapter looks more closely at significant trends and lessons learned from various CGD projects and efforts.

[289] M. Gervais et al., "5 Data Quality Issues and Geographic Knowledge Discovery," *Geographic Data Mining and Knowledge Discovery* (2009): 99–115.

[290] Ibid.

6 Significant Trends/Lessons Learned with Crowdsourced Geospatial Data

Smartphones and Future Geocomputing Trends

Longley, Goodchild, Maguire, and Rhind, in Geographic Information Systems and Science, approach the difficult task of explaining what the future of GIS might look like.[291] According to them, this future world of GIS software deviates from the current desktop-centered GIS paradigm in that the entire system is predicted to be a distributed, web-based architecture using specialized application servers and data servers. The entry point to this future world of GIS is described to be, simply, a standards-compliant web browser. Figure 47, adapted from Longley et al, shows the current desktop-centered paradigm on the left and a future network-based paradigm on the right.

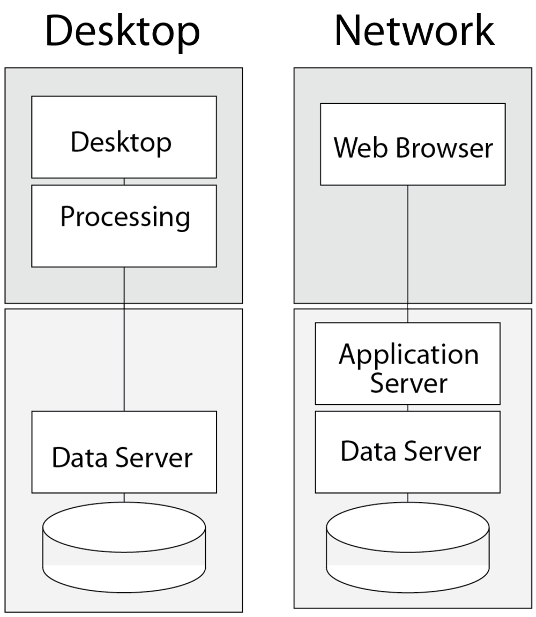


Figure 47. Current Desktop GIS Paradigm and Future Network GIS Paradigm[292]

The current GIS computing environment, according to Longley, will become obsolete as future analysis, processing, and data requests are made through a web browser and executed on a networked application server. Many of the future GIS users will have a variety of thin-client devices whose processing, memory, and operating system will be less important than the presence of software for communicating with an application server located somewhere else on the network.

The current emerging state of networked GIS also involves mobile applications, both web browser-based

[291] Longley et al., *Geographic Information Systems and Science*. p. 181-206

[292] Adapted from Ibid.

and custom apps that are used over a network, primarily on mobile computing devices. Modifying Longley's diagram to emphasize the use of mobile applications and web applications would make it more reflective of the paradigm that is emerging today.

Many of the CGD applications profiled in Chapter 3 use applications developed for mobile devices. Stefanidis et al. describe the interplay of mobile devices, location-aware sensors, and social media as part of a world of ambient geographic information.[293] Many other authors have used the metaphor of citizen sensors to describe a new computing paradigm where mobile computing and user interactivity predominate. Mobile computing and specifically, the smartphone, has become an important element in collecting and using CGD.

Smartphones are rapidly becoming a primary communication device of American consumers. According to a 2012 Pew Research Center study, 46% of American adults own a smartphone, with 74% of those owners using the smartphone to get real-time location information, and 18% using the smartphone to access a geosocial service.[294] For many, these devices are an important means of exchanging email, reading news, accessing entertainment, and maintaining a social network. The smartphone allows its owner to remain in near-constant contact with friends, family and acquaintances. Smartphones also possess powerful sensing platforms used by numerous CGD applications to gather valuable data.[295] Applications use these sensors to capture information such as audio, photo, video, motion, and location via GPS and proximity.

Foursquare,[296] Street Bump,[297] and NoiseTube[298] are examples of current applications that make use of smartphone sensing capabilities while providing valuable information for public services.

[293] Stefanidis, Crooks, and Radzikowski, "Harvesting Ambient Geospatial Information from Social Media Feeds."

[294] The Pew Research Center, "Smartphone Adoption and Usage | Pew Research Center's Internet & American Life Project," 2011, http://www.pewinternet.org/Reports/2011/Smartphones.aspx.

[295] "Sensors in Smartphones | Mobile Device Insight," December 2011, http://mobiledeviceinsight.com/2011/12/sensors-in-smartphones/

[296] "Foursquare."

[297] "Street Bump."

[298] "NoiseTube", 2011, http://www.noisetube.net/

In addition to typical crowdsourcing applications, there are a variety of novel approaches to CGD and smartphones in the domain of accessibility and navigation through unfamiliar environments. Nuernberger,[299] Barbeau et al.,[300] and Rice et al.[301] describe how mobile devices such as smartphones can be extended to improve accessibility by delivering CGD and authoritative geospatial data to the device user. The applications involve obstacle avoidance and reporting, hazard notification, and contextual navigation cues.

Although smartphones provide valuable information, there are inherent security issues exemplified in the case of location sharing applications such as Foursquare that report user locations. Some of these issues relate to the integrity of the smartphone information, while others relate to privacy. Such information could potentially lead to criminal activity against users or inadvertent breaches of privacy. Monmonier provides an excellent starting point for an exploration of the many difficult issues associated with geoprivacy in an age of mobile computing, imaging systems, and surveillance, and notes that this issue will be significant in the years ahead.[302]

Innovation and Cutting Edge Technology

Much of the innovation found with varying CGD platforms can be attributed to either their web-based nature or mobile accessibility. A natural byproduct of this environment is that a large population of potential users and contributors has access to an array of applications. Unlike legacy systems that must remain compatible with older capabilities, CGD applications can be built from the ground-up on the most advanced technology. They evolve rapidly in response to user feedback. This has created opportunities for innovation best captured within an open-source model of development.

---

[299] Andrea Nuernberger, "Presenting Accessibility to Mobility-Impaired Travelers" (UCTC Dissertation, University of California Transportation Center, 2008).

[300] Sean J. Barbeau et al., "Travel Assistance Device: Utilising Global Positioning System-enabled Mobile Phones to Aid Transit Riders with Special Needs," *Intelligent Transport Systems, IET* 4, no. 1 (2010): 12–23.

[301] Rice et al., "Supporting Accessibility for Blind and Vision-impaired People With a Localized Gazetteer and Open Source Geotechnology."

[302] Mark Stephen Monmonier, *Spying with maps: Surveillance Technologies and the Future of Privacy* (Chicago, Illinois: University of Chicago Press, 2002).

The open-source model has increased the utility of these platforms. OSM,[303] for instance, has been developed in a much-more collaborative environment for interested contributors to work. Additionally, significant technologies such as schema-less databases,[304] web-based and mobile editing,[305] search tools such as the image tiling and presentation tools,[306] messaging, and social media have aided in the success of CGD.

As an alternative to traditional geographic information systems, Tomnod has developed an innovative platform optimized for crowdsourcing imagery-based search tasks. The browser-based system tiles imagery into chips, each of which can be analyzed by multiple contributors. The platform includes tools for filtering user input, ranking participants, and identifying hotspots.[307]

Ushahidi[308] is an excellent example of a platform that captures CGD via messaging. Messaging in particular, plays a critical role in the way in which CGD is created and consumed because it relies on more common infrastructure, such as cell phone networks, than more complex, broadband systems. This is particularly important for areas of the world where broadband networks do not exist. For instance, messaging over cell phone networks may be used for disease incident reporting in sub-Saharan Africa, where other infrastructure is largely non-existent.

This non-profit tech company releases open source software aimed at collecting information, visualization, and interactive mapping. Mediums such as this provide the tools necessary to allow a less restrictive flow, better storage, and representation for crowd source information. Syria Tracker[309] is an example of this approach built on the Ushahidi platform.

Organization and Engagement of User Communities

Part of the success and continuing longevity of some CGD applications is due in part to user communities that develop around a project. Goodchild

[303] "OpenStreetMap."

[304] Maxmiliano F. Braga and Fabio N. de Lucena, "Using NoSQL Database to Persist Complex Data Objects," *63ª Reunião Anual Da SBPC* (July 2011).

[305] "Google Map Maker."

[306] "Tomnod: Crowdsource the World."

[307] "Products Overview," *Tomnod*, n.d., http://tomnod.com/products/

[308] "Ushahidi."

[309] "Syria Tracker."

et al.[310] and Zook et al.[311] document the way that CGD projects spontaneously emerge during emergencies. Often, there is no substantial organizing group or mechanism for engaging end-users other than the urgency associated with the disaster or emergency relief effort.

The many volunteers contributing geospatial data during the Santa Barbara wildfires, profiled by Goodchild and Glennon, were not formally organized in any specific way and had no enduring connection or engagement outside the wildfire events. [312]

The truly spontaneous, ephemeral CGD efforts, such as those profiled by Goodchild and Glennon,[313] are sometimes hard to profile due to their highly variable nature.

The comparatively much larger Haitian earthquake response, profiled by Zook,[314] notes a much more organized effort by at least four existing groups, each with an organization and structure. GeoCommons, OSM, CrisisCamp, and Ushahidi could all be reasonably described as organized efforts with structure, leadership, and a user community.

CGD communities may evolve to a longer-lasting, more organized presence. As a noted CGD project with a lengthy history, the organizational structure of OSM is of interest. OSM began as a loosely affiliated group of open-source advocates reacting to the rigid licensing, distribution and copyright controls on geospatial data produced by the Ordnance Survey of Great Britain.[315] After early interest in the project and substantial growth under what could be characterized as a benevolent dictator model centered on Steve Coast, an organization emerged with a meritocratic structure.

Today, the OSM Foundation is a United Kingdom-registered not-for-profit organization that supports the OSM Project.[316] The OSM Foundation has

---

[310] Goodchild and Glennon, “Crowdsourcing Geographic Information for Disaster Response.”

[311] Zook et al., “Volunteered Geographic Information and Crowdsourcing Disaster Relief.”

[312] Goodchild and Glennon, “Crowdsourcing Geographic Information for Disaster Response.”

[313] Ibid.

[314] Zook et al., “Volunteered Geographic Information and Crowdsourcing Disaster Relief.”

[315] Steve Coast, “OpenStreetMap” (Keynote Address, Session 1 presented at the NCGIA Workshop on Volunteered Geographic Information, December 13-17, 2007, UCSB, Santa Barbara, California, December 13, 2007), http://ncgia.ucsb.edu/projects/vgi/docs/present/Coast_openstreetmap-opendata.pdf

[316] “About,” *OpenStreetMap Foundation*, n.d., http://blog.osmfoundation.org/about/

a well-defined structure and organization with a Board of Directors and a number of important working groups, including the Communication Working Group,[317] Data Working Group,[318] Licensing Working Group,[319] Local Chapters Working Group,[320] Operations Working Group,[321] Engineering Working Group,[322] State of the Map Organizing Committee[323] and Strategic Working Group.[324] Each working group has a well-defined role in furthering the mission of OSM. In addition to the OSM Foundation groups, there are user groups spread around the world who organize local OSM meetings and host mapping parties.

The organizational structure of the OSM effort is a model for successful open source application development. This project has been able to produce a high-quality product from a very small core group of highly skilled contributors coupled with a large user community that may often lack formal training in geospatial technologies. The OSM products, produced in a crowdsourced framework, are similar in quality to many commercial products and because of their licensing framework, provide a good source of base map data for many other open-source projects.

Strategies for User Engagement

Crowdsourced geospatial data production represents a new model, as noted in Chapter 2. As with many open source projects, CGD production benefits motivated users and cannot be sufficient without a critical mass of users.

317 "Communication Working Group," *OpenStreetMap Foundation*, November 15, 2011, http://www.osmfoundation.org/wiki/Communication_Working_Group

318 "Data Working Group," *OpenStreetMap Foundation*, April 22, 2012, http://www.osmfoundation.org/wiki/Data_Working_Group

319 "Licensing Working Group," *OpenStreetMap Foundation*, February 15, 2012, http://www.osmfoundation.org/wiki/Licensing_Working_Group

320 "Local Chapters Working Group," *OpenStreetMap Foundation*, November 15, 2011, http://www.osmfoundation.org/wiki/Local_Chapters_Working_Group

321 "Operations Working Group," *OpenStreetMap Foundation*, February 4, 2012, http://www.osmfoundation.org/wiki/Operations_Working_Group

322 "Engineering Working Group," *OpenStreetMap Foundation*, February 4, 2012, http://www.osmfoundation.org/wiki/Engineering_Working_Group

323 "StateoftheMap Organizing Committee," *OpenStreetMap Foundation*, August 24, 2011, http://www.osmfoundation.org/wiki/StateoftheMap_Organizing_Committee

324 "Strategic Working Group," *OpenStreetMap Foundation*, February 4, 2012, http://www.osmfoundation.org/wiki/Strategic_Working_Group

Strategies for user engagement must balance production requirements, quality standards, and the willingness of contributors to participate. Creating a successful method for end-user and contributor engagement may be one of the most important and critical considerations in a VGI application. The success of a CGD project often depends on the motivations of the crowd.

For the emergency response scenarios discussed in Chapter 2 and Chapter 3 of this report, the motivations of the end-users are fairly clear and easy to deduce, and the engagement typically continues as long as the disaster response requires. For emergency response, the motivations of CGD contributors and volunteers reflect the altruistic elements discussed by Goodchild in the initial publication drawing attention to the emerging CGD movement.[325]

For other CGD efforts, such as OSM, the motivation is derived from a continuing resentment over the pricing and licensing practices of the Ordnance Survey.[326] The motivation of CGD contributors is often more complex than altruism or resentment, and may include a desire for self-promotion, a compulsion to fill gaps in areas that lack spatial coverage, and a desire to correct errors.[327]

Although user registration mechanisms in CGD projects are thought to increase quality through accountability, as discussed in Chapter 4, user registration is generally considered to be a disincentive in many open source communities.

A significant engagement method used effectively by OSM and noted as a problem in the search for lost aviator, Steve Fossett,[328] is the communication among community members and the fostering of a community identity through the use of blogs, user discussion forums, educational training material, videos, and methods for facilitating user social connections. The

---

[325] Goodchild, "Citizens as Sensors: The World of Volunteered Geography."

[326] Coast, "OpenStreetMap."

[327] Goodchild, "Citizens as Sensors: The World of Volunteered Geography."

[328] "Steve Fossett Update - National Geographic Adventure Magazine", n.d., http://adventure.nationalgeographic.com/2008/11/special-report/steve-fossett/james-vlahos-text; "Fossett Sought via Google Earth," *BBC*, September 10, 2007, sec. Technology, http://news.bbc.co.uk/2/hi/6987358.stm

necessity for communication was listed as one of many suggestions to improve the Fossett Search crowdsourcing application.[329]

Methods of user engagement that have found particular success in a variety of settings are the use of recognition, rewards, user ratings and user evaluation. Recognition systems provide titles to individuals based on their participation or expertise. On the Old Weather site, individuals are rated as Cadets, Lieutenants, or Captains, based on the number of contributions, providing an incentive for greater participation.

Foursquare ties the user ratings and user history into a system of rewards, such as priority access to events and cash discounts at restaurants. Gasbuddy has a similar system, where registered users who contribute data acquire points that can be redeemed for prizes, as noted in the earlier profile of this application.

Commonly used in computer support forums, user ratings are used to reflect the experience and contributions of the participant. Sometimes, the user status is bestowed directly by a forum manager and may reflect specialized experience or employment status, while in other cases, a user's status may be directly related to the number of contributions, the length of active participation history, or the number of positive assessments from other forum users.

Inevitably, some CGD applications lose participation from end-users and decline, as noted by Goodchild and Glennon for Wikimapia (profiled in Chapter 3), which has been the subject of repeated malicious attacks and subsequent efforts to prevent vandalism.

Goodchild and Glennon note that CGD applications follow a life-cycle, and that Wikimapia is evidence of an application in decline.[330] Other crowdsourced geospatial data applications, like OSM, have not declined in the same fashion because of support from traditional organizations and a transformation from a typical CGD project into a large hybrid project with elements of authoritative control. Because geospatial crowdsourcing is a relatively recent phenomenon, more information of project life-cycles will emerge in the coming years.

---

329 Barbalace, "Internet Search for Steve Fossett Eight Weeks Later."

330 Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

Tailoring Task to Talent

Tailoring tasks to fit individual talents is an important consideration. A potential contributor community may range from experts on the subject matter to novice users and geospatial crowdsourcing projects must insure that the participants have the skills to accomplish the tasks or can learn the skills through training. The success or failure of applications may rely heavily on this particular consideration.

Goodchild[331] contrasts the role of experts in science with amateurs, as reviewed in Chapter 2 of this report. As noted in relation to several citizen science applications such as the Christmas Bird Count,[332] most crowdsourced scientific data produced by amateurs is based on direct observation rather than through analysis or deductive reasoning from observations.[333] "The amateur . . . is limited to engagement in the process of raw observation, and to the inductive rather than deductive role of empiricism."[334]

He notes the role of OSM contributors in making direct observations of position and naming of familiar features: "Mapping of streets and other well-defined features may require simple skills that almost anyone possesses: The ability to use GPS to determine location, and the ability to identify the names and other obvious characteristics of features."[335]

The ability of CGD contributors to observe and identify familiar features is related to the notion of local geographic expertise, which suggests that end-users are likely to contribute CGD in their local neighborhoods, and that this familiar activity space can be thought of as a domain of topical expertise.

In contrast, Goodchild suggests that some mapping projects, such as large mapping of soil types, is a project that clearly falls in the domains of the expert, and would not be a good candidate for crowdsourcing.

[331] Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content."

[332] "Christmas Bird Count."

[333] Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content." p. 12.

[334] Ibid.

[335] Ibid., p. 6.

Crowdsourced geospatial data is ideally directed toward familiar, local, identifiable features that can be easily observed and identified or to tasks that can be addressed by the general public. A review of the applications from Chapter 3 will reinforce this notion about the best domains for CGD.

Malicious and Mischievous Content

Malicious or mischievous content, also referred to as vandalism, is a problem with geospatial crowdsourcing and has been well documented within a number of efforts, including Wikipedia, Wikimapia, OSM, and Foursquare. Vandalism can take the form of mislabeled features, misplaced features, or spoofed coordinates. While not common, in the sense of affecting large quantities of data, vandalism is pervasive among CGD applications.

Users of CGD should be aware of the problem and incorporate methods to identify this content or mitigate the risk associated with intentional disinformation. Producers of CGD need to incorporate methods for rapidly detecting and removing malicious or mischievous data, through automated means or by leveraging Linus's Law through user review.

Clearly, malicious content harms the usefulness of crowdsourced data and erodes the trust in this production technique.

Licensing

Over the last 30 years, intellectual property issues have been prominent in the geospatial community, due to the growth of computer networks, the ease with which digital information is copied and transmitted, and an important Supreme Court case which has had significant consequences for licensing and sharing geospatial data in the United States.

The *Feist v. Rural* Supreme Court case in 1991 changed many aspects of licensing and intellectual property protection in the United States. The case suggested that facts, by themselves are not copyrightable.[336] This idea also suggests that compilations of facts, such as databases and maps, are also not protected by copyright.[337] In response to these events, the National Research Council produced a comprehensive review of the legal mechanisms and arrangements used to share geographic information, and

[336] *Feist Publications ,Inc. V. Rural Telephone Service Co., Inc.*, 499 U.S. 340 (111 S. Ct. 1282 1991).

[337] Dennis S. Karjala, "Copyright in Electronic Maps," *Jurimetrics J.* 35 (1994): 395.

a comprehensive review of the goals, motivations, and benefits to society and government.[338]

In the appendix to the National Research Council report, a number of licensing models and licensing alternatives are reviewed. For CGD, a few licensing models are common, and are generally used to clearly communicate what intellectual property rights are reserved, and which are waived.

The Creative Commons non-profit organization has developed licensing structures used by a number of CGD projects, including OSM, which has used an attribution share-alike version of the Creative Commons license (abbreviate CC BY-SA). This license grants the end-user the ability to copy, distribute, and transmit OSM data and to adapt and commercially reuse the data as long as attribution is preserved and all derivative works and subsequent version of the work preserve the same or a similar license.[339] Several other licenses are available for free use from Creative Commons, with some being more prohibitive and restrictive, barring, for instance, any derivative works and any commercial use (abbreviated CC BY-NC-ND).

Another licensing agreement commonly used in open source and CGD projects is the Open Database License (ODbL), which is a freely distributed product of the Open Data Commons project, run by the Open Knowledge Foundation, a non-profit whose goal is to create a world "in which knowledge is ubiquitous and routine."[340] The Open Database License is similar in some ways to the Creative Commons Share-Alike license but specifically developed for databases. OSM's data is being transitioned from a Creative Commons Attribution-Share-Alike license to an Open Database License, which according to the OSM Foundation is more appropriate for databases. They note that the Creative Commons license was not created for data and the Creative Commons does not recommend using Creative Commons Licenses for data.[341]

---

[338] National Research Council, *Licensing Geographic Data and Services* (Washington, D.C.: The National Academies Press, 2004).

[339] "Attribution-ShareAlike 2.0 Generic (CC BY-SA 2.0)."

[340] "Home," *Open Knowledge Foundation*, 1999, http://okfn.org/

[341] "License/We Are Changing The License," *OpenStreetMap Foundation*, July 31, 2012, http://www.osmfoundation.org/wiki/License/We_Are_Changing_The_License; "License/Why CC BY-SA Is Unsuitable," *OpenStreetMap Foundation*, August 1, 2010, http://www.osmfoundation.org/wiki/License/Why_CC_BY-SA_is_Unsuitable

As discussed in Chapter 2, the USGS has adopted OSM tools for the current incarnation of the CGD-based National Map Corps project, but has not adopted the OSM data, as the licensing is considered too restrictive. Federal agencies in the US are required by law to distribute most unclassified data for the cost of reproduction, while the OSM Foundation uses a variety of Creative Commons and ODbL licenses for data and open-source software.

Another CGD project with an interesting licensing situation worth mentioning is Waze, which was profiled extensively in Chapter 3. Waze is a free GPS-based application that gathers information from users and produces routing, traffic load estimates, and some commercial location information. Waze software is distributed under a GNU General Public License, which is a general use free software license produced by the Free Software Foundation. The base data used in Waze and the crowdsourced content created by end users, however, is not part of this license. Waze's base map data is derived from the US Census Bureau's Tiger Data, which itself is unrestricted. As mentioned in Chapter 3,Waze considered using OSM base data, but the OSM Creative Commons license this would have restricted their commercial use.

Refinements to licensing models for CGD will continue to emerge, and the existence of hybrid projects combining open-source tools and government data is evidence that the difficult legal issues and concerns are being considered and addressed in a way not imagined even 10 years ago.

CGD as Intelligence

Graduate students at George Mason University, when asked to write term papers about crowdsourcing and geospatial data, often cite the role of validation, while noting the general concerns about quality discussed in Chapter 4. A few students have asserted the CGD could be characterized as intelligence, noting that intelligence gathering processes involve many different sources of information, some that provide context, some that provide specific details, and others that provide validation.

In his 1997 memoir, Duane Clarridge describes the development of human intelligence within networks he cultivated as a CIA Officer in Europe, Central America, and Asia.[342] The small bits of intelligence gathered during

[342] Duane R Clarridge, *A spy for all seasons: my life in the CIA* (New York, NY: Scribner, 1997).

brief conversations and through passed messages were often incomplete, imperfect, and in some cases, incorrect. Clarridge's role, as an officer in the clandestine services, was to develop sources, gather intelligence, assess the quality and reliability of the intelligence, and transmit it in a report to Langley. This process of gathering, assessing, and transmitting is reflected directly in CGD and the processes described in this report, which was approved for public release by the CIA.[343]

As with intelligence information, CGD carries a danger of malicious content, including information intended to deceive. CGD may provide a reason to take rapid action even before full verification or assessment has been done (as described by Goodchild and Glennon in their discussion of fire boundary mapping in the Santa Barbara, California area[344]).

CGD has a role in providing verification, and for providing initial estimates in areas where information is sparse. Ultimately, as with intelligence, CGD can be used as one source or perspective from which to construct a larger, more complete picture.

[343] John H. Hedley, "Publications Review Board," *Central Intelligence Agency*, May 8, 2007, https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/docs/v41i3a01p.htm.

[344] Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

7 Summary

The emerging phenomenon of crowdsourced geospatial data (CGD) is an important trend for the geospatial community, as it will have an influence on the future ways that geospatial data is generated, gathered, maintained, and presented. The involvement of end-users in this movement, many of whom are untrained in the geospatial sciences, is an identifying characteristic of this phenomenon. These end-users sometimes referred to as neo-geographers, contribute to geospatial data collection efforts, geospatial application development, and related social media activities.

In 2007, Goodchild coined the term volunteered geographic information (VGI) to describe the largely altruistic activity associated with this neo-geography community. Building on earlier expertise with sensor networks and geointelligence, Stefanidis et al.[345] expanded the boundaries of the VGI phenomenon to include the active harvesting of information from social media and sensor networks, referring to this effort as ambient geographic information (AGI). The collective union of data production activities associated with VGI and AGI is referred to in this report at CGD.

CGD has a few very distinct benefits, as noted by Goodchild and Glennon,[346] Zook,[347] and others. CGD is inexpensive to produce, allowing data to be generated for large areas through volunteer efforts. CGD can also be produced rapidly, as seen in many of the emergency and disaster response efforts profiled in this report. Rapid data production efforts have been facilitated by the availability of high-resolution digital imagery, which can be used as a base layer from which features can be identified, extracted, and digitized.

CGD production tools have also been developed through open-source paradigms, and are widely available and easily adopted. A final and important benefit of CGD is the local geographical expertise of the contributors. Goodchild[348] notes how this expertise is similar to the professional

[345] Stefanidis, Crooks, and Radzikowski, "Harvesting Ambient Geospatial Information from Social Media Feeds."

[346] Goodchild and Glennon, "Crowdsourcing Geographic Information for Disaster Response."

[347] Zook et al., "Volunteered Geographic Information and Crowdsourcing Disaster Relief."

[348] Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content."

expertise manifest in scientific disciplines. Local geographic expertise allows CGD end-users to make data contributions in areas in which they are most familiar, contrasting with authoritative production techniques where no local expertise is typically involved.

CGD production methods, discussed in Chapter 2 of this report, are contrasted with the authoritative data production activities associated with government agencies, large publishers, and non-profit organizations. CGD production is often characterized by a lack of specification and control, as well as an open style of contribution. This production methodology contrasts with the rigid controls, assessments, and specifications present in authoritative geospatial data production. Hybrid geospatial production methods that mix crowdsourcing methods and tools with authoritative methods and tools are being adopted by some government agencies for specific projects, and represent a significant future trend.

A variety of CGD data sources, applications, and activities are profiled in this report, to provide a survey of the large number of emerging efforts in this area. A notable CGD effort, due to its size and success, is OpenStreetMap (OSM), which has a goal of producing a free, editable map of the world. OSM's origins can be traced to the open source movement in the United Kingdom and its reaction to the rigid licensing policies of the Ordnance Survey, the producer of authoritative geospatial data in Great Britain.

Over the past three decades, geographic information systems (GIS) and related technologies have replaced analog geospatial data production methods and paper maps. The traditional ways of assessing the accuracy of data plotted on paper maps is formalized in the National Map Accuracy Standards, which provide guidelines about the acceptable positional errors for well-defined features based on map scale.[349] The National Standard has superseded this for spatial data accuracy, which removes the limitations based on map scale. Many authors have adapted and modified traditional accuracy assessment methods to apply to digital map databases and CGD. The traditional accuracy assessment techniques are summarized in Chapter 4 and discussed in the context of CGD projects, such as OSM.

---

349 "National Geospatial Data Standards - United States National Map Accuracy Standards."

Several additional approaches for quality assessment have been suggested by researchers such as Goodchild,[350] who has been involved in traditional accuracy assessment for decades. These newer approaches include Linus's Law (based on theories of crowdsourcing behavior), hierarchical networks of reviewers, and rules-based triage of geospatial data during production. An important general quality assessment tool for geospatial data is the use of metadata, which is a summary of the content and context of a dataset generated by a data producer. Many excellent metadata standards for geospatial data have been developed, and their use should be promoted within CGD production activities.

Assessing the fitness of CGD for use within a particular project or application involves inspection of metadata (when it exists), visualizations of uncertainty, comparisons of CGD to reference sources (including existing authoritative data for the same area at the same scale), expert reviews and assessments of the CGD (including ratings of content), and a careful assessment of risks and benefits for using CGD and CGD production techniques. In many scenarios with urgent time demands such as disaster response, CGD can provide significant benefits in terms of rapid data generation. In these settings, the benefits may outweigh the risks associated with quality concerns such as the presence of positional error.

This report notes many significant lessons learned and important trends in CGD production and applications. A clear trend, noted both in authoritative production environments and in the crowdsourcing community, is the use of smartphones and other mobile devices. Smartphones, with GPS and other sensors, provide an excellent platform for CGD data collection and data use. Many applications reviewed in Chapter 3 are built to take advantage of the capabilities of smartphones, and their role within general geospatial activities is predicted to increase significantly.

Another important lesson learned from CGD is the importance of the development and engagement of user communities. User engagement is an important part of the success for CGD projects with longevity, such as OSM. Significant attention must be paid to motivating and encouraging users, as CGD efforts rely on a voluntary workforce. Requirements for user-registration and application of complex quality control measures can lead to higher quality contributions, but also to lower participation rates.

---

[350] Goodchild, "Assertion and Authority: The Science of User-Generated Geographic Content."

Elwood et al. provides a valuable critical perspective on the social aspects of CGD and the development of user communities.[351] Another important lesson learned from CGD projects and other crowdsourcing projects is the concern about malicious content and vandalism. The more open a CGD production environment is, the more vulnerable it is to vandalism. Recognizing this problem, groups such as OSM and Wikipedia have developed analytical tools to detect and revert malicious contributions. Those tools are being refined and improved to raise the quality of CGD.

Licensing and intellectual property issues have been raised in the use of CGD, particularly within hybrid environments where CGD and authoritative content are combined. Although not intuitive, the open licensing requirements for CGD may be too restrictive for Government use, as illustrated by the U.S. Geological Survey's decision to forgo OSM for data with no use restrictions. A variety of licensing tools have been developed through the Creative Commons and Open Knowledge Foundation to address the licensing issues present in crowdsourcing applications and crowdsourced data, and the refinement of these licensing tools will continue.

Finally, CGD may best be thought of as intelligence data, rather than traditional map data that is accepted in its entirety. Because CGD frequently lacks the lineage and thorough quality assessment that usually accompanies authoritative geospatial data, there are questions about whether it can be trusted and how reliable it is. At the same time, many experts recognize the tremendous value of CGD, particularly in urgent scenarios. These considerations are very similar in nature to those associated with human intelligence.

A consideration of these lessons learned, and the other material contained in this report, may help individuals and organizations determine whether CGD and CGD-based production techniques are appropriate for their geospatial data production activities.

---

[351] Sarah Elwood, "Volunteered Geographic Information: Key Questions, Concepts and Methods to Guide Emerging Research and Practice," *GeoJournal* 72, no. 3–4 (July 24, 2008): 133–135; Sarah Elwood, "Geographic Information Science: Emerging Research on the Societal Implications of the Geospatial Web," *Progress in Human Geography* 34, no. 3 (2010): 349–357; Sarah Elwood, "Volunteered Geographic Information: Does It Have a Future?" (presented at the AAG Annual Meeting, New York, NY, 2012)

Appendix 1

Acronyms

AGI – Ambient Geographic Information
APAN – All Partners Access Network
API – Application Programming Interface
BAA – Broad Agency Announcement
CGD – Crowdsourced Geospatial Data
CSDGM – Content Standard for Digital Geospatial Metadata
CSV – Comma Separated Values
DARPA – Defense Advanced Research Projects Agency
DHS – Department of Homeland Security
DOD – Department of Defense
EXIF – Exchangeable Image File Format
FGDC – Federal Geographic Data Committee
GFDL – GNU Free Documentation Licenses
GIS – Geographic Information Systems
GNIS – Geographic Names Information Systems
GNS – GEOnet Names Server
GPS – Global Positioning System
GTRI – Global Technology Resources, Inc.
HITS – Human Intelligence Tasks
ICA – International Cartographic Association
IP – Internet Protocol
ISP – Internet Service Provider
JIEDDO – Joint IED Defeat Organization
MIT – Massachusetts Institute of Technology
NGA – National Geospatial-Intelligence Agency
NMAS – National Map Accuracy Standard
NOAA – National Oceanographic and Atmospheric Administration's
NSSDA – National Standard for Spatial Data Accuracy
NYPL – New York Public Library
OCR – Optical Character Recognition
ODbL – Open Database License
OS – Ordnance Survey of Great Britain
OSI – Ordnance Survey Ireland
OSM – OpenStreetMap

OSM CP – OpenStreetMap Collaborative Prototype
PDAs – Personal Digital Assistants
RMSE – Root-Mean-Square Error
SOLAP – Spatial Online Analytical Processing
SOUTHCOM – Southern Command
TED – Twitter Earthquake Detector
UAV – unmanned aerial vehicle
UGC – User Generated Content
URL – Uniform Resource Locator
USGS – United States Geological Survey
VGI – Volunteered Geographic Information
ZIP code – Zone Improvement Plan code

Appendix 2

Table 4. Geospatial crowdsourcing applications summary: Socializing, Georeferencing, and Digitizing

Tasks		Imaging & Georeferencing	Georeferencing	Digitizing		
Examples		Grassroots Mapping	NYPL Map Rectifier	OSM	Google Map Maker	Wikimapia
Data Type		Geographic	Geographic	Geographic	Geographic	Geographic
Primary Interface	Tabular				X	X
Primary Interface	Unstructured					
Primary Interface	Map-based (Source)	USDA NAIP Imagery	OSM	OSM	Google Map	Google Map
Primary Interface	Alternate view	N/A	Google Earth	Cycle Map/ Transport Map/ MapQuest	Satellite	Satellite/ Hybrid/ Terrain/OSM/ Panoramio
Geo-Coverage		Local	Local	Global	Global	Global
Training		Online Video	Online Video	Online Resources Available	Online Resources Available	
Location Input	Type of End-user Reference	Direct Location		Direct Location	Direct Location	Direct Location
Location Input	Point		X	X	X	
Location Input	Line			X	X	X
Location Input	Polygon		X	X	X	X
Place name				X	X	X
Content restrictions		N/A	N/A	N/A	'Appropriate Conduct and Prohibited Actions' policy. Approbation required by Google staff.	N/A
Method for Tracking Contributions		Registration	Registration	Registration	Registration & IP Address	Registration & IP Address
Rating System		No	No	No	No	Yes

Table 5. Geospatial crowdsourcing applications summary: Attributing and Reporting

Tasks		Attributing	Reporting				
Examples		Galaxy Zoo	Louisiana Bucket Brigade	GasBuddy	Street-Bump	Syria Traker	Wikipe-dia
Data Type		Geograph-ic	Non-geographic	Geo-graphic	Geograph-ic & non-geograph-ic	Geo-graphic & non-geograph-ic	Non-geo-graphic
Primary Interface	Tabular		X	X		X	
Primary Interface	Unstruc-tured					X	
Primary Interface	Map-based (Source)	Hubble Telescope imagery	N/A	N/A	Google Map	Google Map	N/A
Primary Interface	Alternate view	N/A	N/A	Proprie-tary	Satellite	Imagery	N/A
Geo-Coverage		Galaxy	Regional	National	Regional	National	Global
Training			Education Provided	None	None	None	None
Location Input	Type of End-user Reference	N/A	Data col-lected from property	Address	Direct Location	N/A	N/A
Location Input	Point			X	X	X	
Location Input	Line				X		
Place name						X	
ZIP code				X			
Content re-strictions		N/A	Require an air sam-pling de-vice to par-ticipate	N/A	Yes	N/A	N/A
Method for Tracking Contri-butions		Registra-tion	Contribu-tors re-quest an air sam-pling de-vice	IP Ad-dress	Required phone applica-tion download	N/A	N/A
Rating System		No	No	Yes	No	No	No

Table 6. Geospatial crowdsourcing applications summary: Searching and Tracking

Tasks		Searching		Tracking	
Examples		Field Expedition: Mongolia - Valley of the Khans Project	DARPA Red Balloon	MapMyWALK	Waze
Data Type		Geographic	Geographic	Geographic	Geographic
Primary Interface	Tabular		X	X	X
	Unstructured	X			
	Map-based (Source)	GeoEye Satellite Imagery	N/A	Google Map	Proprietary
	Alternate view	None	N/A	Satellite/Terrain	N/A
Geo-Coverage		Regional	National		Global
Training		Online video	None	Online instructions available	Instructional videos available
Location Input	Type of End-user Reference	Direct Location	Direct Location	Direct Location	Direct Location
	Point	X		X	X
	Line			X	X
Place name			X	X	X
ZIP code				X	X
Content restrictions		N/A	N/A	None	Terms of Service outline the user submissions limitations. Waze reserves its right to delete any user content they considered inappropriate.
Method for Tracking Contributions		Registration	Registration	Mobile device/ Registration	Registration
Rating System		No	Money reward	No	No

Table 7. Geospatial crowdsourcing applications summary: Transcribing, Validating, and Polling/Survey

Tasks		Transcribing/ Validating	Validating			Polling/ Surveying
Examples		Old Weather	NAVTEQ Map Reporter	Geo-Wiki.org	OSM Inspector	SurveyMapper
Data Type		Geographic	Geographic	Geographic	Geographic	Geographic
Primary Interface	Tabular	X	X	X		X
	Map-based (Source)	Google Map	X	Google Earth	Geofabrik/Mapnik /Open Cycle Map	Google Map
	Alternate view		Satellite/ Hybrid	Satellite		Satellite/ Terrain
Geo-Coverage		Global	Global	Global	Global	Global
Training		Online Instructional video	N/A	Online Video Tutorial	N/A	Online Instructional Video
Location Input	Type of End-user Reference	N/A	Direct Location	Direct Location	N/A	Choice of ZIP code, county, or country.
	Point		X			
	Line		X			
Place name			X	X		X
ZIP code				X		X
Content restrictions		Based on the assumption that wrong contributions can be identified through other 4 right contributions of the same area.	N/A	N/A	Geofabrik-internal data processing	N/A
Method for Tracking Contributions		Registration	IP Address	Registration	N/A	Registration & IP Address
Rating System		Yes	No	Yes	No	No

Table 8. Geospatial crowdsourcing applications summary: Socializing and Sharing

Tasks		Socializing			Sharing	
Examples		Twitter	Flickr	Foursquare	ArcGIS Online	GeoCommons
Data Type		Non-geographic	Non-geographic	Geographic	Geographic	Geographic
Primary Interface	Tabular			X		
	Unstructured	X	X			
	Map-based (Source)	Google Map	NAVTEQ	X	Esri/GEBCO/ NOAA/CHS/National Geographic/DeLorme/NAVTEQ	N/A
	Alternate view		Satellite/ Hybrid		Aerial/Hybrid/Streets/Imagery/ Terrain/ Topographic	N/A
Geo-Coverage		Global	Global		Global	Global
Training		N/A	Online instructions available	No	N/A	Videos, blogs, community forum
Location Input	Type of End-user Reference	Direct Location	Direct Location	Direct Location	N/A	N/A
	Point		X	X	X	
	Line				X	
	Polygon				X	
Place name		X	X	X	X	
ZIP code			X		X	
Content restrictions		N/A	Community guidelines and allowance to report abuse	N/A	N/A	N/A
Method for Tracking Contributions		Registration	Registration	Registration	Registration	Registration

References

39646.png. From Publiclibrary.org, PNG Image, 256 × 256 pixels, December 7, 2011. http://archive.publiclaboratory.org/wcu/2011-12-07-northcarolina-cullowhee-westerncarolinauniversity/tms/16/17625/39646.png.

"A Tour Through GeoCommons." *Geocommons*, n.d. http://geocommons.com/tour.

Aaron. "The Shape of Alpha." *Code: Flickr Developer Blog*, October 30, 2008. http://code.flickr.com/blog/2008/10/30/the-shape-of-alpha/.

"About." *OpenStreetMap Foundation*, n.d. http://blog.osmfoundation.org/about/.

"About Flickr." *Flickr*, n.d. http://www.flickr.com/about/.

Amatriain, Xavier, Josep M. Pujol, and Nuria Oliver. "I Like It... I Like It Not: Evaluating User Ratings Noise in Recommender Systems." In *User Modeling, Adaptation, and Personalization*, edited by Geert-Jan Houben, Gord McCalla, Fabio Pianesi, and Massimo Zancanaro, 5535:247–258. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. http://www.springerlink.com/index/10.1007/978-3-642-02247-0_24.

"Amazon Mechanical Turk - Welcome." *Amazommechanical Turk*, n.d. https://www.mturk.com/mturk/welcome.

Anselm Spoerri. "What Is Popular on Wikipedia and Why?" *First Monday* 12, no. 4 (April 2, 2007). http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1765%E5%AF%86/1645.

"APAN Community", n.d. https://community.apan.org/default.aspx.

"ArcGIS Online." *ArcGIS Online*, n.d. http://www.arcgis.com/home/.

"Attribution-ShareAlike 2.0 Generic (CC BY-SA 2.0)." *Creative Commons*, n.d. http://creativecommons.org/licenses/by-sa/2.0/.

"Australian Newspapers Digitisation Program." *National Library of Australia*, February 17, 2012. http://www.nla.gov.au/ndp/.

Authority, T. V. "Geospatial Positioning Accuracy Standards Part 1: Reporting Methodology". National Aeronautics and Space Administration, 1998. http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/chapter3.

Balloon14. JPEG Image, 2448 × 3264 pixels - Scaled (7%), n.d. http://archive.darpa.mil/networkchallenge/Photos/Balloon14.jpg.

Barbalace, Kenneth. "Internet Search for Steve Fossett Eight Weeks Later." *Yahoo! Groups*, October 31, 2007. http://mx.dir.groups.yahoo.com/group/Rescate/message/8050?var=1.

Barbeau, Sean J., Philip L. Winters, Nevine L. Georggi, Miguel A. Labrador, and Rafael Perez. "Travel Assistance Device: Utilising Global Positioning

System-enabled Mobile Phones to Aid Transit Riders with Special Needs." *Intelligent Transport Systems, IET* 4, no. 1 (2010): 12–23.

Beach, Aaron, Mike Gartrell, and Richard Han. "Solutions to Security and Privacy Issues in Mobile Social Networking." 4:1036–1042. Vancouver, Canada, 2009. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5283078.

"BGN: Domestic Names." *USGS*, April 9, 2012. http://geonames.usgs.gov/domestic/.

Biever, Celeste. "Twitter Mood Maps Reveal Emotional States of America." *NewScientist*, July 21, 2012. http://www.newscientist.com/article/mg20727714.200-twitter-mood-maps-reveal-emotional-states-of-america.html

Bishr, Mohamed, and Werner Kuhn. "Geospatial Information Bottom-Up: A Matter of Trust and Semantics." In *The European Information Society - Leading the Way with Geo-information*, edited by Sara I. Fabrikant and Monica Wachowicz, 365–387. Lecture Notes in Geoinformation and Cartography. Springer-Verlag Berlin Heidelberg, 2007. http://dx.doi.org/10.1007/978-3-540-72385-1_22.

Brimicombe, Allan. *GIS, environmental modelling and engineering*. London; New York: Taylor & Francis, 2003.

"Category:Lists of Mountains." *Wikipedia, the Free Encyclopedia*, February 10, 2012. http://en.wikipedia.org/wiki/Category:Lists_of_mountains.

"Christmas Bird Count." *National Audubon Society Birds*, n.d. http://birds.audubon.org/christmas-bird-count.

Clarridge, Duane R. *A spy for all seasons: my life in the CIA*. New York, NY: Scribner, 1997.

Coast, Steve. "OpenStreetMap". Keynote Address, Session 1 presented at the NCGIA Workshop on Volunteered Geographic Information, December 13-17, 2007, UCSB, Santa Barbara, California, December 13, 2007. http://ncgia.ucsb.edu/projects/vgi/docs/present/Coast_openstreetmap-opendata.pdf.

Coleman, David J., Yola Georgiadou, and Jeff Labonte. "Volunteered Geographic Information: The Nature and Motivation of Produsers." *International Journal of Spatial Data Infrastructures Research* 4, no. 2009 (2009): 332–358.

"Communication Working Group." *OpenStreetMap Foundation*, November 15, 2011. http://www.osmfoundation.org/wiki/Communication_Working_Group.

"Connecting People, Tools and Resources to Support Crisis Response." *CrisisCommons*, 2012. http://crisiscommons.org/.

"CrisisCamp Haiti - Washington DC." *Eventbrite*, 2012. http://crisiscamphaitiwdc.eventbrite.com/.

"DARPA Network Challenge", n.d. http://archive.darpa.mil/networkchallenge/.

"DARPA Network Challenge | FAQ", n.d. http://archive.darpa.mil/networkchallenge/FAQ.aspx.

"Data Working Group." *OpenStreetMap Foundation*, April 22, 2012. http://www.osmfoundation.org/wiki/Data_Working_Group.

Devillers, Rodolphe, Alfred Stein, Yvan Bédard, Nicholas Chrisman, Peter Fisher, and Wenzhong Shi. “Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities.” *Transactions in GIS* 14, no. 4 (August 2010): 387–400.

Devlin, Ian. “Finding Your Position with Geolocation.” Blog. *HTML5 Doctor*, June 14, 2011. http://html5doctor.com/finding-your-position-with-geolocation/.

Division, US Census Bureau Geography. “US Census Bureau TIGER/Line”, n.d. http://www.census.gov/geo/www/tiger/shp.html.

“Dublin Core Metadata Element Set, Version 1.1.” *Dublin Core Metadata Initiative*, June 14, 2012. http://dublincore.org/documents/dces/.

Dunn, Erica H., Charles M. Francis, Peter J. Blancher, Susan R. Drennan, Marshall A. Howe, Denis Lepage, Chandler. S. Robbins, Kenneth V. Rosenberg, John R. Sauer, and Kimberly G. Smith. “Enhancing the Scientific Value of the Christmas Bird Count.” *The Auk* 122, no. 1 (2005): 338–346.

Elwood, Sarah. “Geographic Information Science: Emerging Research on the Societal Implications of the Geospatial Web.” *Progress in Human Geography* 34, no. 3 (2010): 349–357.

Elwood, Sarah, Michael F. Goodchild, and Daniel Sui. “Prospects for VGI Research and the Emerging Fourth Paradigm.” In *Crowdsourcing Geographic Knowledge*, edited by Daniel Sui, Sarah Elwood, and Michael Goodchild, 361–375. Dordrecht: Springer Netherlands, 2013. http://www.springerlink.com/index/10.1007/978-94-007-4587-2_20.

Elwood, Sarah, Michael F. Goodchild, and Daniel Z. Sui. “Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice.” *Annals of the Association of American Geographers* 102, no. 3 (May 2012): 571–590.

Elwood, Sarah. “Volunteered Geographic Information: Does It Have a Future?” presented at the AAG Annual Meeting, New York, NY, 2012.

———. “Volunteered Geographic Information: Key Questions, Concepts and Methods to Guide Emerging Research and Practice.” *GeoJournal* 72, no. 3–4 (July 24, 2008): 133–135.

“Engineering Working Group.” *OpenStreetMap Foundation*, February 4, 2012. http://www.osmfoundation.org/wiki/Engineering_Working_Group.

Feist Publications ,Inc. V. Rural Telephone Service Co., Inc., 499 U.S. 340 (111 S. Ct. 1282 1991).

“Field Expedition: Mongolia.” *National Geographic*, n.d. http://exploration.nationalgeographic.com/.

“Fossett Sought via Google Earth.” *BBC*, September 10, 2007, sec. Technology. http://news.bbc.co.uk/2/hi/6987358.stm.

“Foursquare.” *iTunes Store*, March 27, 2012. http://itunes.apple.com/us/app/foursquare/id306934924?mt=8.

“Foursquare.” *Foursquare*, n.d. https://foursquare.com/.

“Foursquare Addresses Cheating Issue, Frustrates Legit Users.” *CIO Blogs*, April 8, 2012.

http://advice.cio.com/al_sacco/10000/foursquare_addresses_cheating_issue_frustrates_legit_users.
"Free Personal Account." *ArcGIS Online*, n.d. http://www.esri.com/software/arcgis/arcgisonline/features/free-personal-account.html.
"Galaxy Zoo: Classify Galaxies", n.d. http://www.galaxyzoo.org/#/classify
"Galaxy Zoo: Hubble", n.d. http://www.galaxyzoo.org/.
"Galaxy Zoo: Hubble - The Story So Far", n.d. http://www.galaxyzoo.org/#/story
"GasBuddy - Find Cheap Gas Prices." *iTunes Store*, October 14, 2011. http://itunes.apple.com/us/app/gasbuddy-find-cheap-gas-prices/id406719683?mt=8.
"GasBuddy: Save Dollars at the Pump." *Android.AppStorm*, n.d. http://android.appstorm.net/reviews/lifestyle/gasbuddy-save-dollars-at-the-pump/.
"GasBuddy.com - Find Low Gas Prices in the USA and Canada", n.d. http://gasbuddy.com/.
"GeoCommons." *Geocommons*, n.d. http://geocommons.com/.
"Geolocation & IP Location Services With NetAcuity." *Digital Element*, n.d. http://www.digitalelement.com/our_technology/our_technology.html.
"Geospatial Metadata Standards — FGDC Endorsed ISO Metadata Standards." *Federal Geographic Data Committee*, April 25, 2012. http://www.fgdc.gov/metadata/geospatial-metadata-standards#fgdcendorsedisostandards.
"Geospatial Metadata Standards — The Content Standard for Digital Geospatial Metadata (CSDGM)." *Federal Geographic Data Committee*, April 25, 2012. http://www.fgdc.gov/metadata/geospatial-metadata-standards#csdgm.
Gervais, M., Y. Bédard, M. A Levesque, E. Bernier, and R. Devillers. "5 Data Quality Issues and Geographic Knowledge Discovery." *Geographic Data Mining and Knowledge Discovery* (2009): 99–115.
Girres, Jean-François, and Guillaume Touya. "Quality Assessment of the French OpenStreetMap Dataset." *Transactions in GIS* 14, no. 4 (August 2010): 435–459.
"Global Land Cover 2000 | GEM - Global Environment Monitoring", n.d. http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php.
"GlobCover | ESA", n.d. http://due.esrin.esa.int/prjs/prjs68.php.
Goodchild, Michael F. "Assertion and Authority: The Science of User-Generated Geographic Content." In *Proceedings of the Colloquium for Andrew U. Frank´s Birthday*, 3:82–96. Department of Geoinformation and Cartography, Vienna University of Technology, 2008. http://www.tandfonline.com/doi/abs/10.1080/17489720902950374.
———. *Assessing the Value of Neo-Geographic Information: Report on Project Review, 4/27/2012*. Unpublished Project Review Report. Fairfax, VA: George Mason University, April 27, 2012.

———. “Citizens as Sensors: The World of Volunteered Geography.” *GeoJournal* 69, no. 4 (2007): 211–221.
———. “Commentary on Future Trends and Research Objectives: Assessing the Value of Neo-Geographic Information” presented at the AGC-VGI Research Review Meeting, Dr. Matt Rice, chair, George Mason University, Fairfax, VA, April 27, 2012.
———. “NeoGeography and the Nature of Geographic Expertise.” *Journal of Location Based Services* 3, no. 2 (June 2009): 82–96.
Goodchild, Michael F., and J. Alan Glennon. “Crowdsourcing Geographic Information for Disaster Response: a Research Frontier.” *International Journal of Digital Earth* 3, no. 3 (September 2010): 231–241.
Goodchild, Michael F., and Sucharita Gopal. *The Accuracy of spatial databases*. London; New York: Taylor & Francis, 1989.
Goodchild, Michael F. “NeoGeography and the Nature of Geographic Expertise.” *Journal of Location Based Services* 3, no. 2 (2009): 82–96.
Goodchild, Michael F., Phaedon Kyriakidis, Matt Rice, and Philipp Schneider. *Report of the NCGIA Specialist Meeting on Spatial Webs*. Santa Barbara, CA: NCGIA, April 2005.
“Google Docs.” *Google Docs*, n.d. http://www.google.com/google-d-s/forms/.
“Google Map Maker.” *Google*, n.d. http://www.google.com/mapmaker.
“Google Map Maker Terms of Service," n.d. http://www.google.com/mapmaker/mapfiles/s/terms_mapmaker.html
“GPS Accuracy and Limitations.” *Earth Measurement Consulting*, n.d. http://earthmeasurement.com/GPS_accuracy.html.
Gpx_graffiti_vandalism. PNG Image, 1898 × 1130 pixels, n.d. http://wiki.openstreetmap.org/w/images/b/b7/Gpx_graffiti_vandalism.png.
“Grassroots Mapping”, March 9, 2012. http://grassrootsmapping.org/.
Gruver, Adrienne. “Concept Gallery.” Educational resource. *Penn State - College of Earth and Mineral Sciences*, 2012. https://www.e-education.psu.edu/geog486/l8_p5.html.
Guynn, Jessica. “Confessions of a Foursquare Cheater.” *LATimes.com (blog)*, February 16, 2010. http://latimesblogs.latimes.com/technology/2010/02/confessions-of-a-foursquare-cheater.html.
“Haiti Crisis Map - OpenStreetMap NL”, n.d. http://haiti.openstreetmap.nl/.
Haklay, M. “How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets.” *Environment and Planning. B, Planning & Design* 37, no. 4 (2010): 682.
Haklay, Mordechai (Muki), Sofia Basiouka, Vyron Antoniou, and Aamer Ather. “How Many Volunteers Does It Take to Map an Area Well? The Validity of Linus’ Law to Volunteered Geographic Information.” *Cartographic Journal, The* 47, no. 4 (November 1, 2010): 315–322.
Hecht, B., L. Hong, B. Suh, and E. H. Chi. “Tweets from Justin Bieber’s Heart: The Dynamics of the Location Field in User Profiles.” In *Proceedings of*

the 2011 Annual Conference on Human Factors in Computing Systems, 237–246, 2011.
Hedley, John H. "Publications Review Board." *Central Intelligence Agency*, May 8, 2007. https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/docs/v41i3a01p.htm.
Heinis, Thomas, and Gustavo Alonso. "Efficient Lineage Tracking for Scientific Workflows." In *Proceedings of the 2008 ACM SIGMOD International*, 1007–1018. ACM Press, 2008. http://dl.acm.org/citation.cfm?id=1376716.
Hellerstein, Joseph M., and David L. Tennenhouse. "Searching for Jim Gray: a Technical Overview." *Commun. ACM* 54, no. 7 (July 2011): 77–87.
"History of Volunteer Mapping at the USGS." *USGS*, August 2, 2011. http://nationalmap.gov/TheNationalMapCorps/history.html.
"Home." *Open Knowledge Foundation*, 1999. http://okfn.org/.
Howe, Jeff. *Crowdsourcing : why the power of the crowd is driving the future of business*. New York: Crown Business, 2008.
"ICA Commission on GeoVisualization." *ICA ACI*, July 11, 2011. http://geoanalytics.net/ica/.
"Index of Geography Articles." *Wikipedia, the Free Encyclopedia*, November 17, 2011. http://en.wikipedia.org/wiki/Index_of_geography_articles.
Johnson, Peter A., and Renee E. Sieber. "Situating the Adoption of VGI by Government." In *Crowdsourcing Geographic Knowledge*, edited by Daniel Sui, Sarah Elwood, and Michael Goodchild, 65–81. Dordrecht: Springer Netherlands, 2013. http://www.springerlink.com/index/10.1007/978-94-007-4587-2_5.
Karjala, Dennis S. "Copyright in Electronic Maps." *Jurimetrics J.* 35 (1994): 395.
Khan, Tunaggina. "Evaluating the Errors Associated with Zip Code Polygon When Employed for Spatial Analysis". MS Thesis, George Mason University, 2012.
Knutzen, Matt, and Stephen A. Schwarzman. "Drawing on the Past: Enlivening the Study of Historical Geography at Maps.nypl.org." Blog. *NYPL Labs*, February 3, 2010. http://www.nypl.org/blog/2010/02/03/drawing-past-enlivening-study-historical-geography-mapsnyplorg
"LA Bucket Brigade : Air Sample in Norco's Diamond Neighborhood Shows Extreme Levels of Chemicals", n.d. http://www.labucketbrigade.org/article.php?id=803.
"LA Bucket Brigade : Index", n.d. http://www.labucketbrigade.org/.
Laura Kostanski. *To Study the Methods for Recording Unofficial Place Names into Comprehensive Data Sets for Improvement Knowledge Transfer*. Technical Report. Australia: The Winston Churchill Memorial Trust of Australia, 2011. http://www.churchilltrust.com.au/site_media/fellows/2011_Kostanski_Laura.pdf.
Lenders, Vincent, Emmanouil Koukoumidis, Pei Zhang, and Margaret Martonosi. "Location-based Trust for Mobile User-generated Content: Applications, Challenges and Implementations." In *Proceedings of the 9th*

Workshop on Mobile Computing Systems and Applications, 60–64, 2008. http://dl.acm.org/citation.cfm?id=1411775.
"License/We Are Changing The License." *OpenStreetMap Foundation*, July 31, 2012. http://www.osmfoundation.org/wiki/License/We_Are_Changing_The_License.
"License/Why CC BY-SA Is Unsuitable." *OpenStreetMap Foundation*, August 1, 2010. http://www.osmfoundation.org/wiki/License/Why_CC_BY-SA_is_Unsuitable.
"Licensing Working Group." *OpenStreetMap Foundation*, February 15, 2012. http://www.osmfoundation.org/wiki/Licensing_Working_Group.
"List of Cities in Afghanistan." *Wikipedia, the Free Encyclopedia*, April 16, 2012. http://en.wikipedia.org/wiki/List_of_cities_in_Afghanistan.
"Local Chapters Working Group." *OpenStreetMap Foundation*, November 15, 2011. http://www.osmfoundation.org/wiki/Local_Chapters_Working_Group.
Longley, Paul A., Michael F. Goodchild, David J. Maguire, and David W. Rhind. *Geographic Information Systems and Science*. 3rd ed. Hoboken, New Jersey: John WIley & Sons, 2011.
Louisiana Bucket Brigade. *Taking an Air Sample*. From Facebook.com, JPEG Image, 558 × 371 pixels, February 2005. http://sphotos.xx.fbcdn.net/hphotos-ash4/292196_10150641900818963_289341778962_9317550_1210779005_n.jpg.
——. *Untitled*. From Flickr.com, JPEG Image, 540 x 720 pixels, January 1, 1980. http://farm6.staticflickr.com/5041/5244604427_683927c894.jpg.
MacEachren, A. M., C. A. Brewer, and L. W. Pickle. "Visualizing Georeferenced Data: Representing Reliability of Health Statistics." *Environment and Planning A* 30 (1998): 1547–1562.
Map. PNG Image, 819 x 480 pixels, n.d. http://archive.darpa.mil/networkchallenge/BalloonMap.aspx
Mark, David, Andrew Turk, and David Stea. "Progress on Yindjibarndi Ethno-physiography." In *Spatial Information Theory*, edited by Stephan Winter, Matt Duckham, Lars Kulik, and Ben Kuipers, 4736:1–19. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2007. http://dx.doi.org/10.1007/978-3-540-74788-8_1.
Maxmiliano F. Braga, and Fabio N. de Lucena. "Using NoSQL Database to Persist Complex Data Objects." *63ª Reunião Anual Da SBPC* (July 2011).
"MODIS", n.d. http://modis.gsfc.nasa.gov/.
Monmonier, Mark. *How to lie with maps*. 2nd ed. Chicago, Illinois: University of Chicago Press, 1996.
Monmonier, Mark Stephen. *Spying with maps: Surveillance Technologies and the Future of Privacy*. Chicago, Illinois: University of Chicago Press, 2002.

Mooney, P., P. Corcoran, and A. Winstanley. “A Study of Data Representation of Natural Features in Openstreetmap.” In *Proceedings of GIScience*, 150, 2010.
Mooney, P., P. Corcoran, and A.C. Winstanley. “Towards Quality Metrics for Openstreetmap.” In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 514–517, 2010.
Mooney, P., H. Sun, P. Corcoran, and L. Yan. “Citizen Generated Spatial Data and Information: Risks and Opportunities.” In *Proceedings of the 2nd International Conference on Network Engineering and Computer Science (ICNECS 2011)*, 2011.
Mooney, Peter, and Padraig Corcoran. “Characteristics of Heavily Edited Objects in OpenStreetMap.” *Future Internet* 4, no. 1 (2012): 285–305.
———. “Using OSM for LBS – An Analysis of Changes to Attributes of Spatial Objects.” In *Advances in Location-Based Services*, edited by Georg Gartner and Felix Ortag, 165–179. Lecture Notes in Geoinformation and Cartography. Springer Berlin Heidelberg, 2012. http://dx.doi.org/10.1007/978-3-642-24198-7_11.
“Mount Everest.” *Wikipedia, the Free Encyclopedia*, May 15, 2012. http://en.wikipedia.org/wiki/Mount_Everest.
“National Geospatial Data Standards - United States National Map Accuracy Standards.” *USGS*, October 28, 2011. http://nationalmap.gov/standards/nmas.html.
National Research Council. *Licensing Geographic Data and Services*. Washington, D.C.: The National Academies Press, 2004.
“NAVTEQ Corporate - About Us”, n.d. http://corporate.navteq.com/.
“NAVTEQ Map Reporter”, n.d. http://mapreporter.navteq.com/.
“NCGIA Research Initiatives.” *NCGIA*, n.d. http://www.ncgia.ucsb.edu/research/initiatives.html.
“NCGIA Research Initiatives - Initiative 7: Visualizing the Quality of Spatial Information.” *NCGIA*, 1993. http://www.ncgia.ucsb.edu/research/initiatives.html.
“New Twitter - A New Look and Functionality.” *Snap! Websites Journal*, October 2010. http://snapwebsites.info/journal/2010/10/new-twitter-new-look-and-functionality.
“NoiseTube”, 2011. http://www.noisetube.net/.
Nuernberger, Andrea. “Presenting Accessibility to Mobility-Impaired Travelers”. UCTC Dissertation, University of California Transportation Center, 2008.
“NYPL Map Warper: Home”, n.d. http://maps.nypl.org/warper/.
“NYPL Map Warper: Viewing Map 13913”, n.d. http://maps.nypl.org/warper/mapscans/13913.
O’Rourke, Dara, and Gregg P. Macey. “Community Environmental Policing: Assessing New Strategies of Public Participation in Environmental Regulation.” *Journal of Policy Analysis and Management* 22, no. 3 (2003): 383–414 (389).

Oh, Onook, Manish Agrawal, and H. Raghav Rao. “Information Control and Terrorism: Tracking the Mumbai Terrorist Attack Through Twitter.” *Information Systems Frontiers* 13, no. 1 (September 25, 2010): 33–43.

“Old Weather - Our Weather’s Past, the Climate’s Future”, n.d. http://www.oldweather.org/.

“Old Weather Review Interface.” *Vimeo*, n.d. http://vimeo.com/39450854.

“On Foursquare, Cheating, and Claiming Mayorships from Your Couch…” *Foursquare (blog)*, April 7, 2010. http://blog.foursquare.com/2010/04/07/503822143/.

“Online Survey Software - Create Online Surveys.” *Zoomerang*, n.d. http://www.zoomerang.com/.

“Online Survey Software | SurveyGizmo - Advanced Survey Software.” *Surveygizmo*, n.d. http://www.surveygizmo.com/.

“OpenAerialMap”, March 23, 2011. http://openaerialmap.org/Main_Page.

“OpenStreetMap”, n.d. http://www.openstreetmap.org/.

“OpenStreetMap Statistics.” *OpenStreetMap*, n.d. http://www.openstreetmap.org/stats/data_stats.html.

“Operations Working Group.” *OpenStreetMap Foundation*, February 4, 2012. http://www.osmfoundation.org/wiki/Operations_Working_Group.

“Ordnance Survey - OpenStreetMap Wiki”, n.d. http://wiki.openstreetmap.org/wiki/Ordnance_Survey.

“OSM Inspector.” *Geofabrik Tools*, 2011. http://tools.geofabrik.de/osmi/.

“Osmdiff.” *OpenStreetMap Wiki*, June 30, 2011. http://wiki.openstreetmap.org/wiki/Osmdiff.

“OSRM Website.” *OSRM*, 2011. http://map.project-osrm.org/.

“Panoramio - Photos of the World”, n.d. http://www.panoramio.com/.

Paradis, Jeffrey, and Kate Beard. “Visualization of Spatial Data Quality for the Decision Maker: A Data Quality Filter.” *URISA Journal* 6, no. 2 (1994): 25–34.

Pedley, Mary Sponberg. *The commerce of cartography : making and marketing maps in eighteenth-century France and England*. Chicago, Illinois: University of Chicago Press, 2005.

Phillip. “One Million, Six Hundred Thousand New Observations.” Blog. *Old Weather Blog*, July 23, 2012. http://blog.oldweather.org/2012/07/23/one-million-six-hundred-thousand-new-observations/.

“PLOTS Map Knitter”, n.d. http://mapknitter.org/.

“Po Ve Sham – Muki Haklay’s Personal Blog.” Blog. *OpenStreetMap*, March 5, 2012. http://povesham.wordpress.com/tag/openstreetmap/.

“Products Overview.” *Tomnod*, n.d. http://tomnod.com/products/.

“Quality Assurance.” *OpenStreetMap Wiki*, July 18, 2012. http://wiki.openstreetmap.org/wiki/Quality_assurance.

Rana, S., and T. Joliveau. “Neogeography Phenomena-Some Thoughts on It’s Beginning, Future and Related Issues” (n.d.).

Rana, Sanjay, and Thierry Joliveau. “Neogeography Phenomena-Some Thoughts on It’s Beginning, Future and Related Issues” (2007). http://www.ucl.ac.uk/~ucessan/ranajoliveauneogeogpapaper.pdf.

Raymond, Eric S. “Release Early, Release Often.” *The Cathedral and the Bazaar*, August 2, 202AD. http://www.catb.org/~esr/writings/homesteading/cathedral-bazaar/ar01s04.html.

———. *The Cathedral and the Bazaar*. Edited by Tim O’Reilly. 1st ed. Sebastopol, CA, USA: O’Reilly & Associates, Inc., 1999.

Read, Brock. “Can Wikipedia Ever Make the Grade?” *Chronicle of Higher Education* 53, no. 10 (October 27, 2006): 27.

Ren, Mai. “Location Cheating: A Security Challenge to Location-based Social Network Services.” *Computer Science and Engineering: Theses, Dissertations, and Student Research* (December 1, 2011). http://digitalcommons.unl.edu/computerscidiss/31.

Rice, Matthew T. “A Visualization-based Method for Correcting Relative Positional Error Between Topographic Bases”. MS Thesis, Los Alamos National Lab and Brigham Young University, 1998.

Rice, Matthew T., Ahmad O. Aburizaiza, R. Daniel Jacobson, Brandon M. Shore, and Fabiana I. Paez. “Supporting Accessibility for Blind and Vision-impaired People With a Localized Gazetteer and Open Source Geotechnology.” *Transactions in GIS* 16, no. 2 (April 2012): 177–190.

Rice, Matthew T. “Intellectual Property Control for Maps and Geographic Data”. Ph.D. Dissertation, University of California, 2005.

Ricker, W. E. *Computation and Interpretation of Biological Statistics of Fish Populations*. Vol. 191. Fisheries and Marine Service, 1975.

Ruitton-Allinieu, A. M. “Crowdsourcing of Geoinformation: Data Quality and Possible Applications” (2011).

“Script for Reducing the Precision of Nodes.” Wiki. *OpenStreetMap Wiki*, May 15, 2010. http://wiki.openstreetmap.org/wiki/Script_for_reducing_the_precision_of_nodes.

Secretary Committee for Geographical Names of Australia, Australia, and Laura Kostanski. “Crowd-Sourcing Geospatial Information for Government Gazetteer.” In *Tenth United Nations Conference on the Standardization of Geographical Names*. New York, NY, 2012. http://unstats.un.org/unsd/geoinfo/ungegn/docs/10th-uncsgn-docs/crp/E_Conf.101_CRP16_Summary%20paper%20of%20VGI%20for%20UNGEGN.pdf.

“Self-service.” *311 Toronto*, April 25, 2012. https://secure.toronto.ca/webwizard/html/pothole_repair.htm.

“Sensors in Smartphones | Mobile Device Insight”, December 2011. http://mobiledeviceinsight.com/2011/12/sensors-in-smartphones/.

Shirky, Clay. *Cognitive Surplus: Creativity and Generosity in a Connected Age*. Penguin Press HC, The, 2010.

Shore, Brandon M. “VGI Research Review” presented at the AGC-VGI Research Review Meeting, Dr. Matt Rice, chair, George Mason University, Fairfax, VA, April 27, 2012.

Slocum, Terry A., Robert B. McMaster, Fritz C. Kessler, and Hugh H. Howard. *Thematic cartography and geovisualization*. 3rd ed. Indianapolis, Ind.; London: Prentice Hall ; Pearson Education [distributor], 2009.

"Social Weather Mapping with Twitter." *Information Aesthetics*, March 19, 2009. http://infosthetics.com/archives/2009/03/social_weather_mapping.html.

Sood, S.O., J. Antin, and E.F. Churchill. "Profanity Use in Online Communities" (2012). http://research.yahoo.net/files/profanity_chi.pdf.

"Speak To Tweet (speak2tweet) on Twitter", n.d. https://twitter.com/speak2tweet.

Starbird, Kate, and Leysia Palen. "Pass It on?: Retweeting in Mass Emergency." In *Proceedings of the 7th International ISCRAM Conference*. Seattle, WA: International Community on Information Systems for Crisis Response and Management, 2010. http://fsb.cvm.msu.edu/documents/starbirdpaleniscramretweet.pdf.

———. "'Voluntweeters': Self-organizing by Digital Volunteers in Times of Crisis." In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 1071–1080. CHI '11. New York, NY, USA: ACM, 2011. http://doi.acm.org/10.1145/1978942.1979102.

Starbird, Kate. "What 'Crowdsourcing' Obscures: Exposing the Dynamics of Connected Crowd Work During Disaster." In *Collective Intelligence 2012*. Cambridge, MA, 2012. http://arxiv.org/abs/1204.3342.

"StateoftheMap Organizing Committee." *OpenStreetMap Foundation*, August 24, 2011. http://www.osmfoundation.org/wiki/StateoftheMap_Organizing_Committee.

"Statistics." Encyclopedia. *Wikipedia, the Free Encyclopedia*, n.d. http://en.wikipedia.org/wiki/Special:Statistics.

Stefanidis, Anthony, Andrew Crooks, and Jacek Radzikowski. "Harvesting Ambient Geospatial Information from Social Media Feeds." *GeoJournal* (December 4, 2011). http://www.springerlink.com/index/10.1007/s10708-011-9438-2.

"Steve Fossett Update - National Geographic Adventure Magazine", n.d. http://adventure.nationalgeographic.com/2008/11/special-report/steve-fossett/james-vlahos-text.

"Storymill - Tidepool." *Storymill*, 2012. http://storymill.com/tidepool/.

"Strategic Working Group." *OpenStreetMap Foundation*, February 4, 2012. http://www.osmfoundation.org/wiki/Strategic_Working_Group.

"Street Bump", n.d. http://streetbump.org/.

"Submit a New Report." *Syria Tracker*, n.d. https://syriatracker.crowdmap.com/reports/submit.

Sui, Daniel, Sarah Elwood, and Michael F. Goodchild, eds. *Crowdsourcing Geographic Knowledge Volunteered Geographic Information (VGI) in Theory and Practice.* New York, NY: Springer, 2013.

Surowiecki, James. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday, 2004.

"SurveyMonkey: Free Online Survey Software & Questionnaire Tool." *SurveyMonkey*, n.d. http://www.surveymonkey.com/.

"Syria Tracker." *Syria Tracker: Missing, Killed, Arrested, Eyewitness, Report*, n.d. https://syriatracker.crowdmap.com/.

Tang, John C., Manuel Cebrian, Nicklaus A. Giacobe, Hyun-Woo Kim, Taemie Kim, and Douglas "Beaker" Wickert. "Reflecting on the DARPA Red Balloon Challenge." *Communications of the ACM* 54, no. 4 (April 1, 2011): 78–80.

"TDOT 311." *Public Leaf*, n.d. http://www.publicleaf.com/tdot311.

"Tell a Story with Your Photos." *Flickr*, n.d. http://www.flickr.com/tour/#section=tell-a-story.

Thatcher, Jim. "From Volunteered Geographic Information to Volunteered Geographic Services." In *Crowdsourcing Geographic Knowledge*, edited by Daniel Sui, Sarah Elwood, and Michael Goodchild, 161–173. Dordrecht: Springer Netherlands, 2013. http://www.springerlink.com/index/10.1007/978-94-007-4587-2_10.

"The Degree Confluence Project", n.d. http://confluence.org/.

"The Geo-Wiki Project." *Geo-Wiki*, n.d. http://geo-wiki.org/login.php?ReturnUrl=/index.php.

"The National Map Corps." *USGS*, August 2, 2011. http://nationalmap.gov/TheNationalMapCorps/.

The Pew Research Center. "Smartphone Adoption and Usage | Pew Research Center's Internet & American Life Project", 2011. http://www.pewinternet.org/Reports/2011/Smartphones.aspx.

"The PLOTS Archive." *Publiclaboratory.org*, April 22, 2012. http://publiclaboratory.org/archive.

"'This Feature Has Been Locked' - Google Groups", n.d. https://groups.google.com/forum/?fromgroups#!topic/google-mapmaker/tykoUwykD3Q.

"This Is the Home of The National Map Corps." *USGS*, May 9, 2012. https://my.usgs.gov/confluence/display/nationalmapcorps/Home.

"Timeline of Updating Process." *Waze*, February 12, 2012. http://www.waze.com/wiki/index.php/Timeline_of_updating_process.

"Tomnod: Crowdsource the World." *Tomnod: Crowdsource the World*, n.d. http://tomnod.com/.

"Transcribe Bentham", n.d. http://www.ucl.ac.uk/transcribe-bentham/.

Turner, Andrew. *Introduction to Neogeography*. O'Reilly Media, Inc., 2006.

Twaroch, Florian A., Christopher B. Jones, and Alia I. Abdelmoty. "Acquisition of a Vernacular Place Names from Web Sources." In *Weaving Services and People on the World Wide Web*, edited by Irwin King and Ricardo Baeza-Yates, 195–213. Springer Berlin Heidelberg, 2008. http://dl.acm.org/citation.cfm?id=1367808.

"Twitter: About." *Twitter*, April 7, 2012. https://twitter.com/about.

"User Guide: Philosophy - Wikimapia", n.d. http://wikimapia.org/wiki/User_Guide:_Philosophy.
"UserActivity." *OpenStreetMap Wiki*, August 24, 2011. http://wiki.openstreetmap.org/wiki/UserActivity.
"Ushahidi." *Ushahidi*, n.d. http://ushahidi.com/.
"Vandalism." *OpenStreetMap Wiki*, July 10, 2012. http://wiki.openstreetmap.org/wiki/Vandalism.
Veregin, Howard, and Gary Hunter. "Data Quality Measurement and Assessment." Educational resource. *The NCGIA Core Curriculum in GIScience*, 1998. http://www.ncgia.ucsb.edu/giscc/units/u100/u100_f.html.
Vieweg, S., A. L. Hughes, K. Starbird, and L. Palen. "Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness." In *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, 1079–1088, 2010.
"Waze - Social Traffic & Navigation App", n.d. http://www.waze.com/.
"Waze: Way to Go", n.d. http://www.waze.com/faq/#6.
"Welcome to Flickr!" *Flickr*, n.d. http://www.flickr.com/.
"Welcome to SurveyMapper." *SurveyMapper*, n.d. http://www.surveymapper.com/.
"Welcome to Twitter." *Twitter*, n.d. https://twitter.com/.
West_Harrisburg. JPEG Image, 1716 × 954 pixels, n.d. http://wiki.openstreetmap.org/w/images/a/a2/West_Harrisburg.jpg.
"What Is SKYWARN?" *NWS SKYWARN*, May 10, 2011. http://www.nws.noaa.gov/skywarn/.
"Wikimapia - Let's Describe the Whole World!", n.d. http://wikimapia.org/.
"Wikipedia", n.d. http://www.wikipedia.org/.
"Wikipedia:Wikipedians." Encyclopedia. *Wikipedia, the Free Encyclopedia*, August 16, 2012. http://en.wikipedia.org/wiki/Wikipedia:Wikipedians#Number_of_editors.
"Wikipedia.org Site Info." *Alexa*, n.d. http://www.alexa.com/siteinfo/wikipedia.org.
Wolf, Eric B., Greg D. Matthews, Kevin McNinch, and Barbara S. Poore. *OpenStreetMap Collaborative Prototype, Phase One*. Open--File Report. U.S. Geological Survey, Reston, VA: U.S. Department of the Interior, U.S. Geological Survey, 2011. http://pubs.usgs.gov/of/2011/1136/pdf/OF11-1136.pdf.
Zielstra, D., and A. Zipf. "Quantitative Studies on the Data Quality of OpenStreetMap in Germany." In *Proceedings GIScience*, 20–26, 2010.
Zook, Matthew, Mark Graham, Taylor Shelton, and Sean Gorman. "Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake." *World Medical & Health Policy* 2, no. 2 (July 21, 2010): 6–32.
Zook, Matthew. "Volunteered Geographic Information: Does It Have a Future?" presented at the AAG Annual Meeting, New York, NY, 2012.